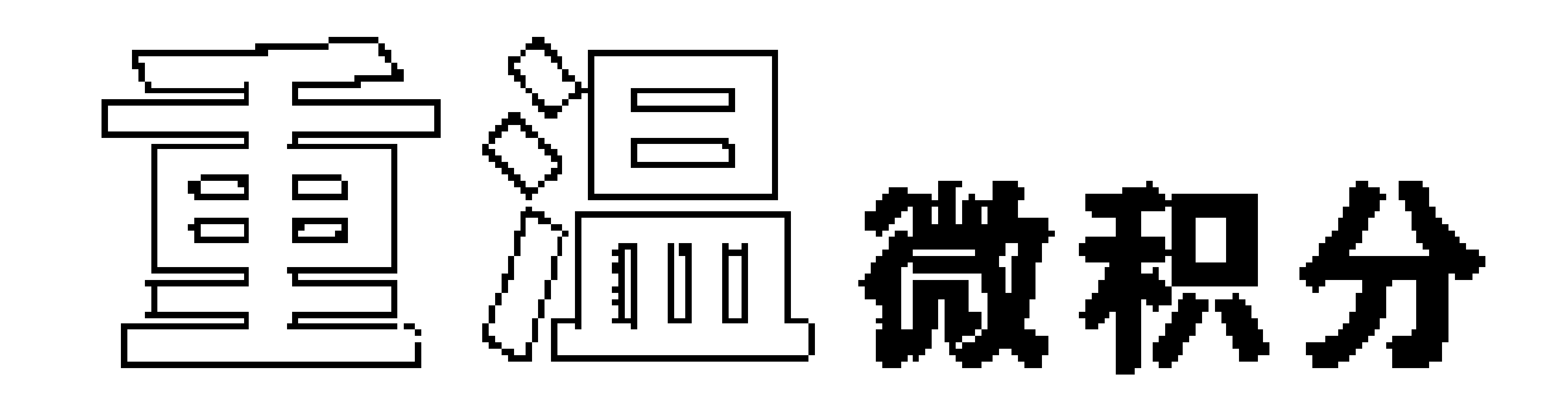



齐民友

高等教育出版社

# 郑 重 声 明



**策划编辑**　李　蕊

**责任编辑**　郭思旭

**封面设计**　王凌波

**责任绘图**　宗小梅

**版式设计**　马静如

**责任校对**　尤　静

**责任印制**

## 内容提要

本书根据作者多年来为各种不同程度的大学生和研究生讲课及讨论班上报告的内容整理而成。第一章对极限理论的发展作了历史的回顾。以下六章分别讨论函数、微分学、积分学、傅里叶分析、实分析与点集拓扑学基础以及微分流形理论。每一章都强调有关理论的基本问题、基本理论和基本方法的历史的背景，其与物理科学的内在联系，其现代的发展与陈述方式特别是它与其他数学分支的关系。同时对一些数学和物理学中重要的而学生常常不了解的问题作了阐述。因此，它涉及了除微积分以外的许多数学分支：主要有实和复分析、微分方程、泛函分析、变分法和拓扑学的某些部分。同样对经典物理学——牛顿力学和电磁学作了较深入的讨论。其目的则是引导学生去重新审视和整理自己已学过的数学知识，并为学习新的数学知识——例如数学物理做准备。

本书适合于已学过微积分的基本知识的大学生和研究生进一步自学更现代的数学之用，也可以作为讨论班的材料。本书还适合需要较多数学的各专业的人员以及高等学校教师参考之用。

**图书在版编目（CIP）数据**

重温微积分/齐民友.—北京：高等教育出版社，2003.12

ISBN 7－04－012931－0

Ⅰ.重...　Ⅱ.齐...　Ⅲ.微积分　Ⅳ.O172

中国版本图书馆 CIP 数据核字（2003）第 091947 号

| | | | |
|---|---|---|---|
| 出版发行 | 高等教育出版社 | 购书热线 | 010－64054588 |
| 社　　址 | 北京市西城区德外大街 4 号 | 免费咨询 | 800－810－0598 |
| 邮政编码 | 100011 | 网　　址 | http://www.hep.edu.cn |
| 总　　机 | 010－82028899 | | http://www.hep.com.cn |
| 经　　销 | 新华书店北京发行所 | | |
| 印　　刷 | | | |
| 开　　本 | 787×1092　1/16 | 版　　次 | 年　月第 1 版 |
| 印　　张 | 35 | 印　　次 | 年　月第　次印刷 |
| 字　　数 | 860 000 | 定　　价 | 39.60 元 |

等等);测度论、概率论与泛函分析;李群与微分几何;代数方面可以讲一点表示论(群论基础知识已经在一二年级讲了)。此外可以有一些选修课。这些建议自然有一些随意性,而且有一些是日常"闲谈"讲的话,算不得教学计划。但是他们中大多数人是这个看法。问他们为什么,回答是:"数学就是统一的,应该让学生有一个统一的认识"。而且他们对我国"分工过细"常表示很不习惯。我在写这本书时,也试着在这方面作一些努力。但是这样就要回答一个问题:例如本书涉及一些所谓实分析知识,是否可以用它来取代实分析课程?我想是不行的。因为既然已分划出这样一个分支,总有自己的道理:有自己的问题,自己的方法,想要掌握它,必须系统地下功夫。所以我不幻想本书能取代其它的书,而只是告诉读者,曲径可以通幽。读者在学了微积分以后,就很容易进入某个领域,并且试着引导他去看一下这个领域的概貌。至于读者是否要进入某个领域,那是读者的事:要看有没有时间,有没有需要,而最重要的是有没有兴趣。我所能做的事无非是趁读者游兴正浓之时,告诉他还有哪些地方值得漫游,值得探胜乃至探险。一是提高他的游兴,二是如果他真的能去了,不至于过于生疏。根据同样的思想,本书可以只读其一章,可以从任一章开始,这与一般教科书也不太相同。这是第二点。

第三点是我们近年来常感到一个问题:可否让学生尽快地进入科学的前沿?另外,在研究生的教学或讨论班中又时常发现有不少内容大学本科就可以懂或者应该懂,而研究生有些弄不清的问题很可能是忘记了他在学微积分时实际上已经接触过这个问题,或者是因为变了一个样子就不认识了。科学发展太快,而学生学习年限不可能再延长。而且将来至少有一部分学生要在没有学校和老师的条件下自己去钻研新的科学知识。所以在大学中就应该为他们准备条件。结论是应该早一点让学生接触新成果。问题是能否做到。上面讲的试验班的经验给了我们一些启发:一方面是要认真地提取出新成果的实质并与传统的教学内容结合起来。否则,老是一本书一本书念下去绝不是办法。其次是老师们要"想开"一点:是不是认真教好书是教师的职责;是不是认真读好书是学生的职责。二者虽有密切关系却不是一件事。教学的某一部分内容时常并非最基本的,学生是否愿意在这里花工夫应该可以自己选择。即令听了不甚了然,至少没有什么坏处。我的一位老师当年就告诉过我,年轻时代思想最敏锐,即令一时不懂,将来再接触到会有似曾相识之感,对自己大有好处。这段话是千真万确的。这本书的内容有相当一部分是我曾以种种形式对学生们讲过的,效果不一定比讲传统的教材更差。走什么样的路,甚至数学系的学生将来是不是一定会以数学为生,这确实是学生们自己的事。"想开"了这一点,就可以放开种种疑虑,反而可以集中思想把书写得更清楚易读一些。

然而对读者们也有一个要求,即要求他们曾读过一本比较认真的传统的微积分教本。例如同济大学应用数学教研室编的《高等数学》:此书立论平正,平易近人,易教易学,作为进一步学习的出发点是够用的。我对现在的教材绝大部分是肯定的,因为它们帮助读者了解一门科学的基本内容。没有这一点,"重温"二字从何谈起?温故而知新,是希望读者能由此再进一步。这是本书写作的目的。因为假设读者有了这样的基础,我就可以略过许多材料不讲,可以假设读者自己会证明——至少知道——许多定理,特别是本书可以不按顺序,全由读者的兴趣从任何一章开始。这本书不是教本,完全不必要全都读懂。有许多内容则是因自己学力所限没有涉及,特别是有关计算机和数值计算的问题。但是关于本书材料的选择,还有几点深感不尽妥当。一是本书没有一章讲常微分方程,而这是应该有的。这是因为 2001 年冬我曾应邀在天津大学和南开大学

为当时还未打算写这本书，那本讲义的内容与本书有些互相牵扯。再重写则头脑里有了一个比较固定的框架，不太好办了，只好等以后有机会再说了。二是本书涉及不少线性代数知识，而基本的又是对偶空间的知识。目前的线性代数教材从一般的线性空间与线性算子角度来讨论的比较少，本书又未能如同处理微积分的古典内容那样细致地处理它，估计会使读者感到困难。如果能在第六、七两章适当补充，效果会更好些。

可是，正如一些数学家一再强调的，数学是“算”懂的，而不是“看”懂的。毫无疑问，这是当前数学教学一大弱点，而且似有加剧之势。本书对所讲的内容力求给以比较清晰的陈述，并给出证明。力求避免“显而易见”，“不难知道”之类的说法。但是对如何有助于读者的计算能力总是考虑不够。一些结论的证明因为与常见的书不同，难免有错，盼读者随时指正。

本书写作首先要感谢老朋友康宏逵老师，他关于数学基础、逻辑以及对微积分的发展的一般看法，多年来与我常共同切磋。本书中引用拉卡托斯的论述更是得益于他。刘伟安老师对大学生(不止是数学专业)的数学教学与我经常讨论。本书中关于向量与张量的处理得他之益甚多。如何给大学生讲一点广义函数更是我们近来交往的主题。田谷基老师对量子力学有兴趣，本书中关于量子力学知识的介绍实际上是我们共同工作的结果。王维克老师从写作本书意念的萌生，一直给予支持和鼓励。全书的打印，刘、田二位老师花了不少精力，对于他们以及许多没有提到的同志，谨致诚挚的谢意。但是我不能不提到责任编辑郭思旭同志。我们是二十多年的老朋友了，我们能合作到退休以后，想起来确有些激动。同样，我向从事编辑和校对的同志深切致谢。

齐民友

2003年3月

# 目　　录

# 第一章　变量的数学
## ——从直观与思辨到成熟的数学科学

读者都读过了一本通常的微积分教本，这样就会知道这是一门很有用的科学，尽管从这类教材中他很少能见到新的例子.再说，一门科学是否很有用也不是只靠几个例子能说明的.读者们会懂得了微积分中有许多解决问题的方法.如果不是遇到了很难的题目，或很细致的定理，微积分不是一门很难念的课程，而应该是很生动的.但是很多读者都对微积分的数学方法很不以为然.具体地说就是很不习惯 $\varepsilon-\delta$ 之类的语言，很不满意于对许多概念的过分仔细的分析.所以，本书打算这样开始：首先从历史发展的轨迹说明微积分何以有这样不"友好"的"界面"(users-unfriendly-interface)？但是本章并不是一个比较全面的微积分学历史的介绍，所以许多重要的人和事都没有讲.我们的目的只在于说明何以会有 $\varepsilon-\delta$ 这样令常人望而生畏的东西.说明这正是科学进步的结果，是数学科学区别于其它科学最明显的特点.本章结束以后，我们将再就微积分的若干主要领域介绍这种语言与方法是如何更有效地表述了微积分的主要思想，如何更有效地刻画了宇宙的规律，同时在这个过程中深化了自己，发展了自己(包括自己的语言与方法).

"初等数学是常量的数学，高等数学是变量的数学".这是老生常谈了，而且大体也是正确的.变量的数学在刻画自然界，乃至人类社会生活中取得了何等辉煌的胜利，这都是人所共知的了.但是什么是变量的数学？因为将"变"的概念引入数学又引起了何等深刻的变化？乃至于什么是"变"或"变量"？这些问题是值得我们去进一步思考的.我们将从历史的发展来看一下，这些问题是如何进入数学家的视野的.当代数学的一个最主要的起源地是希腊.希腊文明的所谓古典时期(即公元前 6 世纪至前 3 世纪)，数学就已经形成了一个独立的学科.在那时，数学与哲学的关系是密不可分的，希腊人对许多数学问题的处理还有浓厚的思辨色彩.然而就是这样，关于变量和变化的数学问题已开始孕育了.简略地回溯一下这段历史，有助于我们去体会为什么微积分会有今天这样的形状，为什么我们不得不绞尽脑汁来对付 $\varepsilon-\delta$.也会体会到，两千多年前提出的问题至今仍未完全"解决".但是，这样做，就不得不进入哲学的领域.这是我们力不能及的.所以，我们只能作一些浅显的介绍，而建议有兴趣的读者去读一些比较专门的著作.我们愿向读者推荐两本书：

M.Kline，Mathematics-The Loss of Certainty，Oxford University Press，1980.有中文译本：李宏魁译，数学：确定性的丧失，湖南科学技术出版社，1997.

T. Dantzig，Number-The Language of Science，George Allen & Unwin Ltd，1938.有中文译本：苏仲湘译，数：科学的语言，上海教育出版社，2000.

这两本书都是严肃的科普著作，不过要请读者注意不要以为读了它们就是读了哲学了.

实验,就可以发现宇宙的规律,而不必求助于超自然的力量.希腊人较少受宗教的束缚,敢于反对传统,敢于反对教条的权威.对于他们,科学的任务就在于发现宇宙的规律.数学是科学的一部分(在希腊时期,几乎是科学的大部分),其任务也就是发现宇宙的规律.而且由希腊人,而伽利略,而牛顿……又总是认为宇宙的规律乃是数学的规律.这当然不是说,数学不为各个时代的经济发展的要求服务,不为技术服务,而是说,数学还有更深刻的任务,即探讨宇宙(原来主要是指自然界,但近几个世纪以来又越来越多地涉及人的社会生活)的规律.正因为它是科学,所以它又能更好地适应技术和经济发展的要求.对这个问题我们也不打算多作讨论,而只是指出,看到了这条主线就更容易理解微积分为什么会成为今天这样子.

数学作为一种演绎推理的科学,即数学中一切结论都应该用逻辑方法加以证明,按照罗素的说明,这个"原则"或者说是"规定",是从毕达哥拉斯开始的.其后经由希腊时代的许多学派(主要是柏拉图学派)形成一个不可动摇的传统,而在欧几里得的《几何原本》中得到了完美的体现.毕达哥拉斯的数学又带有一种神秘的色彩,即他认为宇宙的本质是数(正整数,其实有理数也在其内,因为整数 $n$ 是以 1 为单位,重复 $n$——一个正整数——次而得,而有理数$\frac{n}{m}$则是以$\frac{1}{m}$为新单位重复 $n$ 次而得,这个新单位重复 $m$——又是一个正整数——次又可得原单位 1).既然如此,任何一个线段都应该由若干(正整数)个"单元"构成.这当然是原子论在数学中的反映.所以,任意两条线段 $l_1, l_2$都由相应的正整数 $m_1, m_2$个单元构成.单元是公共的单位,若两条线段 $l_1, l_2$有公共的单位,使它们分别由 $m_1, m_2$个单位构成($m_1, m_2$是正整数),我们称这种情况为 $l_1, l_2$"可公度".但是其后毕达哥拉斯一位弟子正是利用了他所证明了的著名的毕达哥拉斯定理(中国称为勾股定理)发现,正方形的边长(设为 1)与对角线长(用我们现在的记号是$\sqrt{2}$)是"不可公度的",即不可能找到任何的单元使 1 与$\sqrt{2}$各含 $m$ 与 $n$(均为正整数)个单元,亦即不论 $m$ 与 $n$ 是什么样的正整数均不能使$\frac{n}{m}=\sqrt{2}$.用我们现在的语言来说,即$\sqrt{2}$是无理数.这个证明应该是读者们所熟知的.按数学史家的研究,是欧几里得或其弟子们提出的.我们这里只想提到一点,即当时就已经用反证法来证明这个结论.

这是希腊数学中出现的一次危机.它的实质在于:有理数本质上只可用于刻画离散的对象,而几何图形如线段等等本质上却是连续的.但是什么叫连续,我们暂时还只能直觉地去掌握它.至少现在我们知道,用离散性的有理数不能刻画连续性的对象.希腊人还没有掌握无理数的理论,那是两千多年以后的事了.而希腊人对数学的要求又是很严格的,因此,可以说希腊人回避了数而专注于几何.他们研究量而回避数.例如线段之长是量,面积大小也是量.对于两个量例如线段 $l_1, l_2$之长可以讨论它们的"比"(ratio).那么什么是比呢?在《几何原本》,卷Ⅴ中,给出了以下的"定义":"比是两个同类量之间的一种大小关系".今天看来,这实在算不上是一个"定义",但《几何原本》中确实是这样说的.从现代数学的要求看来,《几何原本》在严格性方面是大有问题的,所以两个量之比是不是一个数是有问题的.但是紧接着讲到比例(即两个比之相等)原书上有一个定义 5 就很有意思了.因为照抄原书不太好懂,所以不妨用现代的语言说明如下:设有四个量,例如四条线段 $l_1, l_2, l_3, l_4$,如果对任意两个整数 $m$ 和 $n$,凡当 $ml_1 \gtreqless nl_2$时,必有 $ml_3 \gtreqless nl_4$就说 $l_1$与 $l_2$之比等于 $l_3$与 $l_4$之比.这里要注意 $ml_1 > nl_2$并不是数的比较,而是长度的比较:"把 $l_1$首尾相连,延长 $m$ 倍"大于"把 $l_2$首尾相连延长 $n$ 倍",所以定义 5 是完全没有问题.如果我们

承认量就是数，所以 $l_1$ 与 $l_2$ 之比就是 $l_1/l_2$，则这个定义讲的是：若由 $ml_1 \gtreqless nl_2$ 必可得 $ml_3 \gtreqless nl_4$，就说 $\frac{l_1}{l_2}=\frac{l_3}{l_4}$. 这个定义隐含地承认了“比”虽然不知是不是数，但是可以比较大小. 如果我们把 $ml_1 \gtreqless nl_2$ 定义为 $l_1$ 与 $l_2$ 之比 $\gtreqless \frac{n}{m}$ $\left(\frac{n}{m}\text{确实是毕达哥拉斯也承认的数}\right)$，则这个定义可以改述为，“若 ($l_1$ 比 $l_2$) $\gtreqless$ (某数)，则 ($l_3$ 比 $l_4$) 也 $\gtreqless$ (同数)，这时就说 ($l_1$，与 $l_2$ 之比) = ($l_3$ 与 $l_4$ 之比)”. 我们都知道，数（不论是否有理数）可以比较大小，希腊人则知道量可以比较大小. 但是矛盾在于并不是所有的量都可以用数（实际上指有理数）来表示，例如单位正方形的对角线就不行. 如果我们再多加一句话：能比较大小的就是数，则量就与数完全统一了. 希腊人就是转不过这个弯子来. 是不是希腊人笨？能创造如此辉煌的数学的民族会笨吗？转不过弯子来的症结何在？因为希腊人认为数毕竟要由“单元”构成，正如物体是原子构成的那样. 原子是不可分的，所以单元就是不可分量. 原子是一切整体的组成部分，所以单元作为一切整体的部分必小于一切数. 这就是《几何原本》开宗明义的卷Ⅰ公理 5：“整体大于部分”. 所以，这个“单元”就是“无穷小”. 但是什么是无穷小呢？希腊人不仅是“不知道”，而且还可以证明“无穷小”不存在（下面我们要讲这件事）. 所以希腊人无论如何也转不过这个弯子来. 这个弯子一直到 19 世纪六七十年代才转过来. 例如戴德金（J. W. R. Dedekind，1831—1916），用有理数的分割作为实数的定义，不妨说就是定义：凡能比较大小并作四则运算的东西就叫做实数（我们这样说还不尽符合戴德金的原意. 详见本书第六章）. 那么“无穷小”是什么呢？戴德金不需要无穷小，而是在实数理论基础上建立了极限理论，然后定义无穷小其实是极限为 0 的变量. 这个弯子绕得太大了，不仅希腊人绕不过来，伽利略、牛顿都没有绕过来. 而且从《几何原本》到戴德金，花了两千多年时间，所以如果说《几何原本》中的比例理论（一般认为应归功于欧多克萨斯（Eudoxus，公元前 3 世纪））就是实数理论的前身，那又太有点“戏说”的味道. 但是说从事后看来，希腊数学中已包含了其萌芽，恐怕是事实.

芝诺（Zeno of Elea，约公元前 5 世纪）和他的四个悖论更把关于数的“单元”（原子、不可分量、无穷小）所蕴含的深刻的矛盾突出出来了. 芝诺其人生平不明，著作也没有流传下来. 我们现在所见的来自亚里士多德的《物理学》. 实际上，亚里士多德是企图解决芝诺所提出的矛盾的. 关于芝诺的四个悖论说法很多，下面我们叙述的是罗素在《西方的智慧》（B. Russell，Wisdom of the West，MacDonald & Co，Ltd. 1959）一书中的说法. 该书中文译者是马家驹、贺霖，世界知识出版社 1992 出版，48～50 页.

芝诺的悖论是针对毕达哥拉斯的. 前两个悖论针对的是一条直线是由无限多个单元构成的这样的论点. 第一个悖论是阿基里斯（Achilles，希腊神话中的善跑者）永远追不上乌龟. 阿基里斯的速度比乌龟快 10 倍，但是他让乌龟先走了 100 米. 当阿基里斯追了这 100 米时，乌龟又向前走了 10 米，当阿基里斯追上这 10 米时，乌龟又向前走了 1 米. 如此进行下去直至无穷，阿基里斯是永远追不上乌龟的. 驳倒这一点并不难，因为若阿基里斯的速度为 $v$，则乌龟的速度是 $v/10$. 当阿基里斯追到乌龟的出发点时，他用了时间 $100/v=T$. 而在这个时间里，乌龟向前走了 $\frac{vT}{10}$. 依此类推，阿基里斯所用的时间是

$$T+\frac{T}{10}+\frac{T}{10^2}+\cdots = T\sum_{n=0}^{\infty}\frac{1}{10^n} = \frac{10}{9}T.$$

这里涉及了无穷级数. 很可惜，希腊人不懂得无穷级数，因此他们只能用思辨的方法来解释这些

悖论.

第二个悖论是所谓“两分法”(dichotomy).“阿基里斯”是用相对运动来反对直线是由无限多个单元组成这一论断的,“两分法”则说运动是不可能的,并以此反驳上述论断:若一个物体要沿某一直线走完一定的路程,它必先经过其中点;但在达到中点以前,先要到达 1/4 点;在到达 1/4 点以前先要到达 1/8 点……仿此进行下去,该物体必须在有限时间内走完无限多个区段.而这是不可能的.因此,运动不可能.当然,用现代的数学知识来解释这个悖论也不困难.只要承认时间是无限可分的,则相应于空间的无限多个区段,也有时间的无限多个区段.而这些时间区段的长度成为一个几何级数,公比是 1/2,因而无限多个时间区段长之总和可以是有限的.这就解释了两分法悖论.

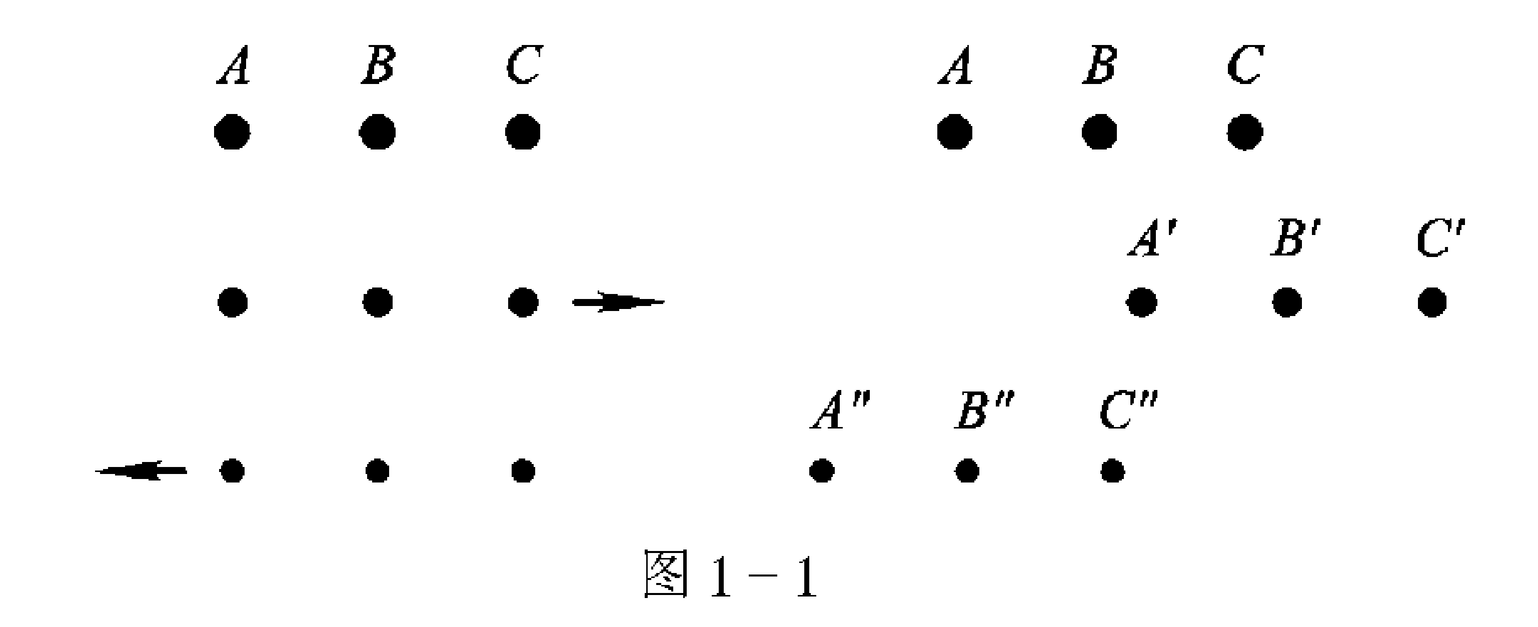



图 1-1

另两个悖论则是说:假设直线由无限多个单元组成,固然会引起如上的悖论,假设它是由有限多个单元组成,同样也会产生悖论.这里的悖论又是一组两个,第一个用相对运动,另一个则用一个物体的运动.第三个悖论称为“运动场”(stade).假设空间的一个线段是有限多个单元组成的,一个时间区段则由有限多个瞬间(即时间的单元)组成.若有三排点 $A,B,C$ 如图 1-1.第一排不动;第二排向右移动,由于单元总数为有限,所以点的移动距离只能是有限个单元,所用的时间区段也只包含有限多个瞬间.所以我们可以假设每个瞬间移动一个单元.第三排则向左移动一个单元.总之点的移动只能是在一个瞬间(即一个时间单元)内,移动一个点(即一个空间单元).但是如图 1,原来的 $A$ 点在一个瞬间以后,向右移动到 $A'$,向左移动到 $A''$,相距两个点的距离.于是 $A''$这一排点相对于$A'$所经过的点数两倍于相对于$A$ 所经过的点数,即一瞬间走了两点.但是按单元总数为有限的假设,一瞬间只能走一个点.由此可见,运动是不可能 的.这一段话是直接从亚里士多德《物理学》一书中引用的.

第四个悖论是飞箭不动.飞行中的箭在每一个瞬间都“占有本身大小相等的位置一定的时间”.因此在这个瞬间,箭是不动的.而时间是由有限多个瞬时组成的.因此,在整个时间中,飞箭都是不动的.矛盾的产生在于假设了时间与空间都是由有限个单元组成的.用我们现在的观点来看,问题在于芝诺讲到了速度,如果用 $x(t)$表示物体的位置,则在两个不同瞬间 $t_1$与 $t_2$,有速度就是假设了

$$v=\frac{x(t_2)-x(t_1)}{t_2-t_1}\neq 0.$$

因此 $x(t_2)-x(t_1)\neq 0$,所以就必然有运动.芝诺说,因为飞箭在每一瞬间都有一定位置,这就意味着飞箭在这一瞬间是静止的,这个论断在逻辑上就没有根据.上面我们说的其实就是导数的概念.飞箭在每一瞬间都有固定的位置,即 $x(t)$有确定的值,这并不是静止.要静止,至少需要

$x'(t)=0$. 当然，希腊人没有这样的概念，芝诺的悖论有违于现代的科学，但是芝诺用这些悖论来反驳时间空间是由“单元”构成的，确实是击中要害的.

这些悖论表明，我们直觉中以为没有什么困难的概念如运动、连续，变化等等，如果深入地进行逻辑的分析，就会发现我们并不真正懂得它们. 但是因为芝诺提出这些问题时完全是从思辨出发的，因此，对芝诺的反驳也必然是思辨的. 其中最重要的是亚里士多德. 他把无限分成两类：实无限(actual infinity)和潜无限(potential infinity). 前者是指某一时刻“一下子”就掌握到无限，后者则是在一个时间的过程中才能逐步展开、实现的无限. 例如一个线段可以无限地分割，但在任一个具体的时刻都不能无穷尽分割到头(因此不存在毕达哥拉斯所谓单元)；1，2，3，4，…可以无限地数下去，时间可以不断流逝，这些都是潜无限. 亚里士多德认为，芝诺的种种悖论都是针对的实无限. 因此，亚里士多德认为如果只限承认潜无限，则不但不会有悖论，而且正是抓住了现实世界的本质，芝诺的困难也都迎刃而解. 不仅如此，亚里士多德还指出一条直线的“连续性”就表现在：如果把它分成左右两段，则左段的终点正是右段的起点，二者融合为一；时间也是这样，0 时 0 分 0 秒既是昨天的最终一瞬，也是今天的第一瞬，二者也融合为一. 但是数就不是这样；以整数 1，2，3，4，5，6，…而言，如果在 4 处分开，则 4 是左段的最末一数，5 是右段的第一数，其间还差了许多. 即令是有理数也是一样；$\sqrt{2}$正是 1，1.4，1.41，1.414，…与 2，1.5，1.42，1.415，…之间的空隙. 数的集合究竟有没有连续性，一直没有解决. 希腊人没有系统的无理数理论，这个理论出现在 19 世纪中后叶戴德金(Dedekind)与康托尔(Cantor)之手，正是回答了亚里士多德的问题. 欧几里得一直只注重几何，而尽量回避数，原因大概在此.

芝诺的悖论在希腊哲学中影响极大，而后人对芝诺悖论的诠释、评论也众说纷纭，有人说甚至有过于对待《圣经》. 以上我们只从数学的发展来看，就立刻看出，这里的关键在于“单元”究竟是什么？或者说，“无穷小”、“不可分量”究竟是什么？今天回顾历史，实在不能责怪希腊人回答不了这些问题. 在世界上的一切古老文明中，只有希腊人如此深刻地提出了问题. 一直到 19 世纪中后期的柯西(Cauchy)、魏尔斯特拉斯(Weierstrass)才给出了我们现在满意的回答. 而在当时，至少是在柏拉图、欧几里得这里，数学家们持极为严谨的态度，实际上是“回避”无穷小. 而他们所采用的方法，实际上是走了亚里士多德潜无限的路，而且实际上开辟了我们今天的极限理论的道路.

不少人认为希腊人这样的研究数学实在是脱离实际和烦琐哲学，而且播下了今天的数学之脱离实际的“罪恶”的种子. 对此实难苟同. 因为运动、连续、变化不是我们每天都看得见的么？不是宇宙本性的一部分么？因此不可避免地成为数学家研究的对象. 何况，在希腊人看来，数和其它赋形的事物如日月星辰、山川草木的区别其实不大. 只不过数学的方法与当时已经存在的科学：天文学、光学、静力学及至生理学等等不同，它不能依靠观察与实验，而只有逻辑推理，这当然又与毕达哥拉斯开创的传统相联系. 看一看当时一些杰出的数学家的研究，就会发现他们实际上都是沿着亚里士多德潜无限的思想前进的. 也许人们说这是亚里士多德的哲学思想的“指导作用”，但是我们没有任何证据来证明这种说法. 宁可说，当时希腊人的认识水平就只达到了这一步.

我们要举出的杰出数学家之一是欧多克萨斯(Eudoxus). 他生于公元前 408 年左右，比亚里士多德早了好几十年，所以没有“接受”亚里士多德哲学“指导”的好福气. 他是希腊古典时代最伟大的数学家之一，可是其生平与著作俱无流传. 一般认为，欧几里得《几何原本》关于比例的理论，

关于穷竭法(早期的积分法)等等都是他的贡献.欧多克萨斯的工作只讲量(如长度、角度、面积、体积……)而不讲数,这是为了避免毕达哥拉斯关于无理数的困难.他应用了一个原理,后来人们认为应该看作是一个公理,因为阿基米德(Archimedes)也用过它,所以称为 Archimedes-Eudoxus 公理,见《几何原本》卷Ⅹ,命题1(这里和以下凡引用《几何原本》处,均可在李文林主编《数学珍宝——历史文献精选》(科学出版社,1998 年出版)中找到).用我们现代的语言来说,它是:设有两个量 $a>b>0$,从 $a$ 中除去其一半以上,再从其余除去其中一半以上,如此进行下去,经过充分多次以后,所余必小于 $b$.这个原理极为重要,因为它排斥无穷小量(不可分量)的存在:为简单起见,每次均取 $a$ 的 $\lambda$ 倍,$\frac{1}{2}\leqslant\lambda<1$,即 $\lambda a$,于是所余为 $(1-\lambda)a$,$n$ 次以后所余当为 $(1-\lambda)^n a$,这个命题说,只要 $n$ 充分大,必有 $(1-\lambda)^n a<b$.其实,$\lambda$ 不必 $\geqslant\frac{1}{2}$,$\lambda>0$ 即可,欧多克萨斯取 $\lambda\geqslant\frac{1}{2}$是为了下面穷竭法所需.如果有一个量是无穷小量,即是小于一切量而又不是 0 的量,则令上面的 $b$ 为无穷小量,而取 $a$ 为单位量 1,则一定有一个 $n$,使 $(1-\lambda)^n<b$,所以 $b$ 一定大于某个非 0 的量 $(1-\lambda)^n$ 而不会小于一切非 0 量.所以,无穷小量是不存在的.这一段证明显然有亚里士多德的潜无限的味道.

欧多克萨斯然后用它来求圆的面积.严格地说,他并没有求出圆的面积,而是证明了下面的

**定理** 圆与圆之比等于其直径上的正方形之比.(《几何原本》,卷Ⅻ,2.)

我们仍要比较仔细地用现代语言来叙述《几何原本》中的证明,因为这就是早期的积分学.设有两圆,其直径分别为 $d$ 与 $d'$,面积相应地为 $S$ 与 $S'$,则待证明的就是下面的比例式:

$$S:S'=d^2:d'^2. \tag{1}$$

或

$$S/d^2=S'/d'^2. \tag{2}$$

我们当然知道,这个公比是 $\pi/4$,但是注意到希腊人连$\sqrt{2}$都不承认,又怎会承认 $\pi$ 呢?所以欧多克萨斯干脆不讲这个公比是什么.但到了阿基米德则进一步计算了 $\pi$ 的近似值,他分别用圆的内接正 96 边形和外切正 96 边形算出 $\pi=3.14103$ 和 $\pi=3.14271$,因此,$\pi$ 的真值在二者之间.

这个定理的证明分为两部分,第一部分是证明圆可以内接正多边形去“穷竭”,即只要多边形的边数充分大,内接正多边形的面积就与圆面积近似到任意要求的程度,因此,这个方法称为穷竭法.不过这个名词并非来自希腊人,而是来自 17 世纪的欧洲.证明如下,记圆面积为 $S$,内接正 $n$ 边形面积为 $P_n$,再作一个各边平行于此 $n$ 边形的外切正 $n$ 边形,记其面积为 $Q_n$.很明显

$$P_n<S<Q_n. \tag{3}$$

而且,只要 $n\geqslant4$,很容易看到 $P_n\geqslant\frac{1}{2}Q_n$.(《几何原本》中是用 $n=4$ 即用内接外切正方形来完成这个证明的.)$S-P_n$ 由 $n$ 个相同的弓形构成.如果把内接正 $n$ 边形边数加一倍,即设 $AB$ 为其一边,取$\overset{\frown}{AB}$之平分点 $C$,联结 $AC$,$BC$,用这两个边代替 $AB$ 即得正 $2n$ 边形,其面积为 $P_{2n}$.$S-P_{2n}$是 $2n$ 个相同的弓形.今证这 $2n$ 个弓形面积小于$\frac{1}{2}(S-P_n)$.为此,只看一个边 $AB$,并作相应的矩形▭$ABED$ 如图 1-2,则$\triangle ABC=\frac{1}{2}$▭$ABED$,而(弓形 $ABFCGA$)的面积小于 $2\triangle ABC$.

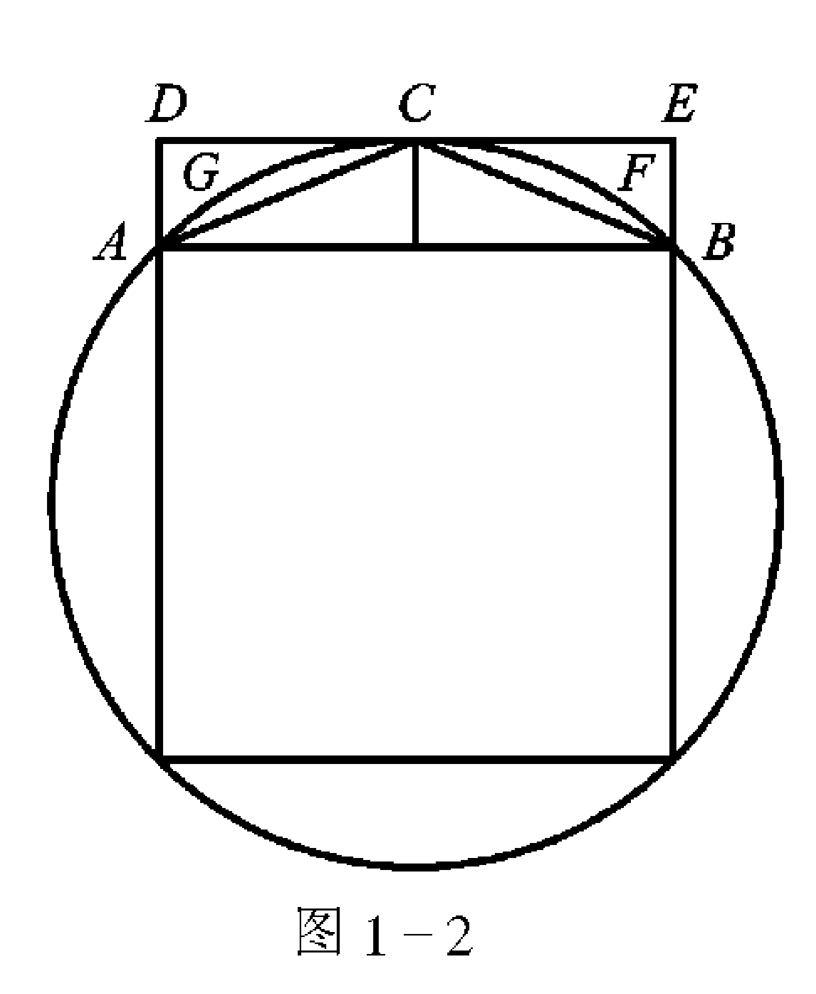


图 1－2

如果我们把内接正多边形边数增加一倍，则 $S-P_{2n}$ 与 $S-P_n$ 的关系是从(弓形 $ABFCGA$)中减去 $\triangle ABC$，而把原来的一个弓形换成两个小弓形，其弦各为 $\widehat{AC}$ 与 $\widehat{BC}$. 但被减去的 $\triangle ABC > \frac{1}{2}$(弓形 $ABFCGA$)，所以这一个弓形被减去了“一大半”. 每一个弓形如此，$n$ 个弓形之总和当然也如此. 所以随着每一次把内接正多边形的边数加一倍，$S-P_n$ 就会被减去“一大半”. 仿此进行下去，利用卷Ⅹ命题 1，即知，只要 $n$ 取得足够大，$S-P_n$(其实就是 $|S-P_n|$)就会变得“要多小有多小”.

至此，读者们一定会说，由此推知，对任意 $\varepsilon>0$ 必定有 $N$ 存在，使当 $n>N$ 时，$|S-P_n|<\varepsilon$，亦即 $\lim\limits_{n\to\infty} P_n = S$. 但是不要太急，不要忘记希腊人还没有极限的明确概念. 而他们对数学严格性的要求又使他们不能用没有定义过的东西.(不过，他们实际上并没有真正做到这一点，例如什么是圆面积，他们就只说它是一个量，而不是数，但是什么是量，什么是量的比都没有下定义. 在 20 世纪初希尔伯特又写了一本《几何基础》(科学出版社有中译本)，这样才达到了 20 世纪数学严格性的要求，但是已把《几何原本》改得面目全非了). 不使用极限，使希腊人绕了一个大弯——求助于反证法.

证明的第二部分是求证(1)式，这里他们用了反证法. 设(1)不成立，则一定可以找到比例第四项 $S''$ 使

$$S : S'' = d^2 : d'^2. \tag{4}$$

关于比例第四项是否一定存在，《几何原本》又是默认而未证明的. 如果 $S''<S'$，则取 $n$ 充分大，如上所述可以使 $S'$ 的内接正 $n$ 边形 $P'_n$ 与 $S'$ 之差任意小，从而小于 $S'-S''$，即

$$|P'_n - S'| < |S' - S''|,$$

于是有

$$S' > P'_n > S''. \tag{5}$$

在《几何原本》中紧紧放在这个定理前还有一个命题：圆内接相似多边形之比等于圆直径平方之比，这又是很容易证明的. 故

$$P_n : P'_n = d^2 : d'^2. \tag{6}$$

再由(4)，有

$$P_n : P'_n = S : S''.$$

因 $P_n<S$，故 $P'_n<S''$，但这与(5)矛盾.

如果设 $S''>S'$，也同样会出现矛盾.

定理证毕.

我们不厌其烦地引证了二千多年前的文献，是为了与我们现在都懂得的微积分作一比较. 我们甚至不妨说，19 世纪中叶，柯西和魏尔斯特拉斯正是复兴了欧多克萨斯的穷竭法，才完成了微积分基础的奠定. 首先，他们正式给出了极限的定义，这是欧多克萨斯没有做到的；有了极限，他们就以 $P_n$ 的极限作为圆面积的定义，这也是欧多克萨斯没有想到的；有了这一切，他们就再也用不着绕反证法的大弯子了. 他们定义无穷小即是极限为 0 的变量，从而“真正”地排除了毕达哥拉

斯“单元”以及后来的不可分量的“痕迹”.这只是在欧多克萨斯的基础上走了更明确的一步.他们也是宣告了亚里士多德潜无限思想的“胜利”，把笼罩着的思辨色彩“清除”了.这一点恐怕他们自己也没有想到.千真万确的是，他们把变量、极限等等归结为不等式，确是得了欧多克萨斯的真传.可惜的是，正是这一点成了 $\varepsilon-\delta$ 的罪名，因为它们使人望而生畏了，这却是变量的数学的精华.

不过，凡事有利必有弊.《几何原本》在数学上的严格性成了人们的“模范”以后，也必然会对人的直觉、感觉、不严格的推理有所排斥.而这种排斥时常也把许多极为生动的，通向真理的其它道路封锁了.以亚里士多德而言，潜无限的思想被推到了主导的地位，实无限的思想难道就一无是处了吗？否，历史证明不是这样.《几何原本》一个最明显的缺点是轻视数学的实际应用.这不是说希腊数学整个都是轻视应用的.即以欧多克萨斯而言，他不仅是一位数学家，还是一个天文学家(他曾经提出一个关于天体运动的模型)，他还是一个医生，一个地理学家.另一位更著名的大人物是阿基米德(Archimedes，287—212，B.C.).他是亚历山大时期(或称希腊化时期)希腊数学和科学的代表人物.他较晚于欧几里得，但二人大异其趣.他不仅在数学，而且在力学、机械学、光学等各方面都有大贡献.他还是一位伟大的发明家，而欧几里得主要贡献则是完成了《几何原本》(其中不少是他的弟子们的功绩)，有点像孔夫子说的“述而不作”，而阿基米德的创造天才却是光彩夺目的.因为我们的目的只是讨论微积分的思想与方法的源起，所以现在就不来讨论希腊数学各个时期特点的比较及其功过了.好在到了公元前 2 世纪至前 1 世纪，罗马人征服了希腊.出现了罗马帝国，希腊数学也退到历史的后台去了.

千年沉睡以后，当数学科学再次醒来，已经开始进入资本主义时代了.社会经济的发展要求变了.到了 16 和 17 世纪，人类面临新的数学问题.其中一些重大问题是：求一般几何形体的长度，面积，体积；求曲线的切线；求运动的速度与加速度；求函数的极大与极小等等.我们的读者一眼就可以看出来，当时人类需要的新科学就是微积分.这些问题一方面有明确的实用背景，但又常与哲学上的论争，与人类争取从天主教神学以及亚里士多德的教条下获得思想解放的斗争紧密联系.伽利略在比萨斜塔上的实验(但是科学史的研究表明，他并没有做这个实验，而是用斜面的实验反驳了亚里士多德)、围绕日心说的斗争是大家都熟知的了.新时代呼唤巨人，他们是笛卡儿、哥白尼、伽利略、开普勒直到牛顿和莱布尼茨.下面是一些例子.

第一个例子是关于积分学的.那个时代人们努力的重点是发展希腊人的穷竭法，但是他们又丢了穷竭法中的精华：用复杂的反证法绕过缺少极限理论带来的困难，把问题归结于不等式，从而为 19 世纪的 $\varepsilon-\delta$ 铺平了道路.他们反而又回到了不可分量(亦即毕达哥拉斯的单元)，不过更大胆，而且几乎没有多少玄妙的思辨色彩了.开普勒(Johannes Kepler，1571—1630)有一本书名叫《测量酒桶体积的新科学》，顾名思义，其中会讲到面积和体积.在开普勒看来，圆是无穷多个顶角无穷小的扇形堆成的，球是无穷多个互相平行的厚度无穷小的圆柱形堆成的.伽利略也是这样，例如他认为若一物体(应为质点)作等加速运动，则在 $t=0$ 到 $t=T$ 这段时间走过的距离将由 $\triangle OAB$ 之面积来表示.因为，若在时刻 $t$，速度为 $A'B'$，伽利略将它乘以无穷小的时间间隔，$\triangle OAB$ 就是由无穷多个直线段 $A'B'$ 组成的，因此他得出结论：距离由

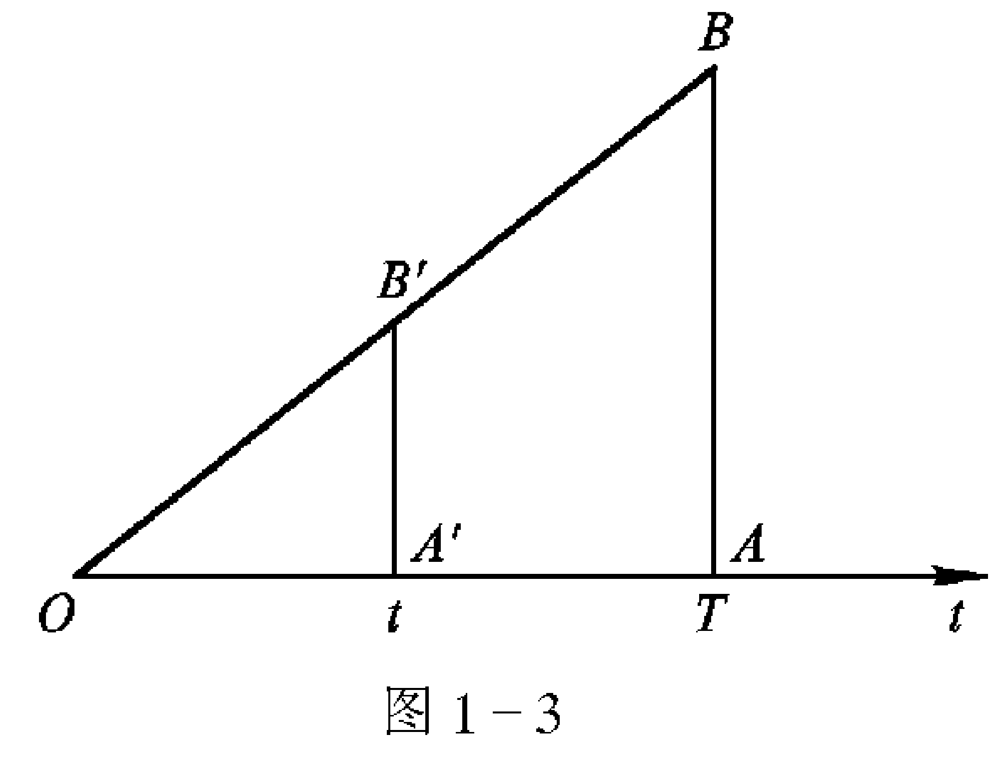


图 1－3

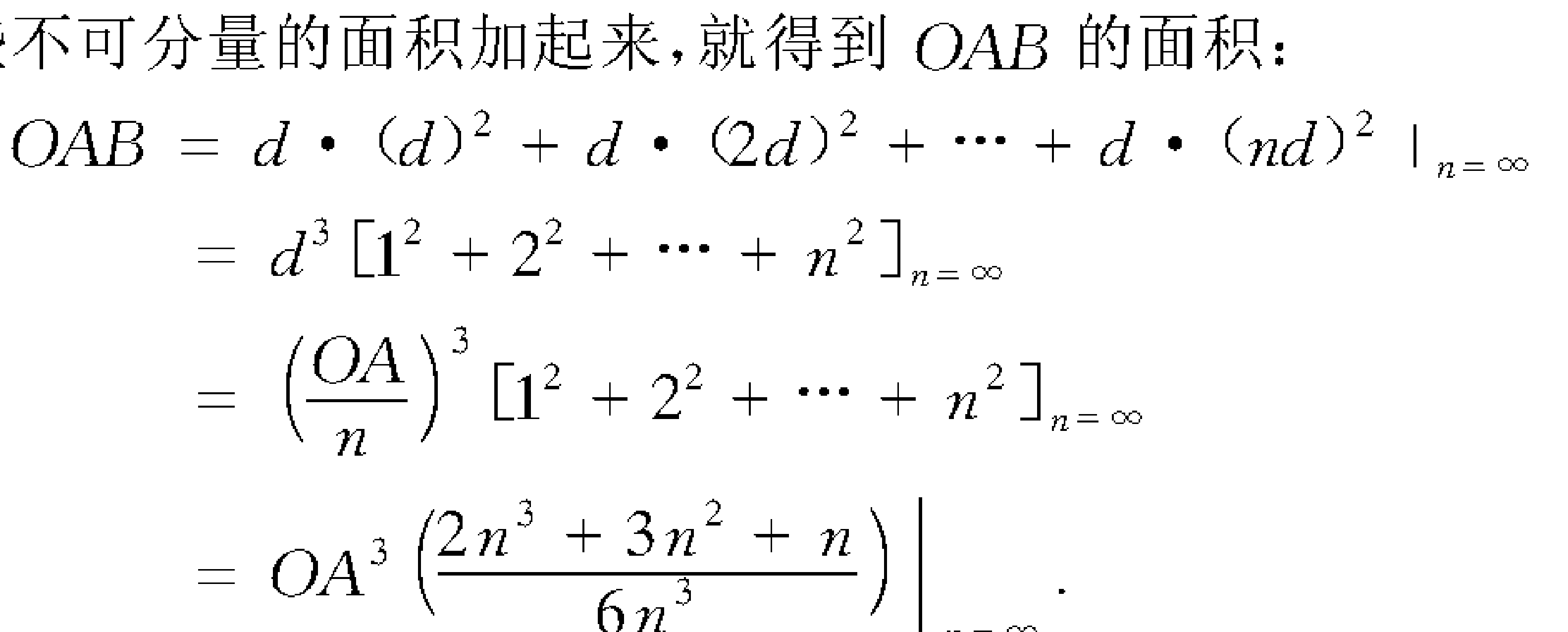
$\triangle OAB$ 之面积来表示(图 1-3).

这样看来,伽利略和开普勒在面积体积问题上是一样的,采用了不可分量的观点.这种观点在伽利略的学生卡瓦列里(Bonaventura Cavalieri, 1598—1647)的著作里表现得最系统.简单地说,他认为线段是由点组成的,正如链子是由珠子穿成的一样;面是由线组成的,正如布是由纱织成的一样;立体是由平面组成的,正如书是一页一页的纸叠起来的一样.点、线、面就是相应的不可分量,它们在一个或几个维数上是无穷小.

所以,如果说欧多克萨斯是在千方百计地回避无穷小,而为此费尽心思利用了 Archimedes-Eudoxus 公理,利用了不等式,利用了反证法,这一个时代的数学家却是硬着头皮往前闯.虽说不严格,却得到了许多正确的结果,这样不会招致人们的批评吗?卡瓦列里的回答是绝妙的.他说,我的目的只是为了改善穷竭法:我的方法确实不严格,但是我的结果很有用,有用就行.严格不严格那是哲学家的事,别的几何学家不是和我一样不严格吗?

首先,它确实改善了穷竭法.上面我们讲到的希腊的穷竭法太依赖于特殊的几何曲线(例如圆)的特殊性质,现在却有了一定的程式,例如要计算一段抛物线 $y=x^2$ 下的面积 $OAB$ 时(图 1-4),我们把 $OA$ 分成 $n$ 等分,每一段的长记为 $d$,于是相应的纵坐标是 $d^2, (2d)^2, (3d)^2, \cdots$.如果把 $n$ 取为无穷大,那么图上用实线画的区域都成了直线段,我们用虚线把它们补成矩形的.如果 $n$ 是无穷大,这些矩形和实线所成图形是没有区别的,它们就是不可分量.现在把这些不可分量的面积加起来,就得到 $OAB$ 的面积:

图 1-4



$$\begin{aligned}OAB &= d\cdot(d)^2 + d\cdot(2d)^2 + \cdots + d\cdot(nd)^2 \big|_{n=\infty}\\ &= d^3\left[1^2+2^2+\cdots+n^2\right]_{n=\infty}\\ &= \left(\frac{OA}{n}\right)^3\left[1^2+2^2+\cdots+n^2\right]_{n=\infty}\\ &= OA^3\left(\frac{2n^3+3n^2+n}{6n^3}\right)\Bigg|_{n=\infty}.\end{aligned}$$

这里只用了一个很初等的代数公式

$$1^1+2^2+\cdots+n^2=\frac{1}{6}n(n+1)(2n+1).$$

还有 $\left.\frac{1}{n}\right|_{n=\infty}=0$.前式并不难证,后面的 $\left.\frac{1}{n}\right|_{n=\infty}=0$ 按卡瓦列里的说法,交给哲学家去管好了.不管怎样,许多人都可以按这个程式去计算积分了.如果把严格性暂置一旁,这确实是很大的进步.

第二个例子是关于速度.速度和加速度这些运动学概念的研究在当时的重要性我们已在上面讲过.如果运动轨迹用 $x=x(t)$ 来表示,则大家知道 $\frac{\Delta x}{\Delta t}$ 表示平均速度.现在要求瞬时速度,可是又不能令 $\Delta t=0$,因为那样一来就连运动也没有了(牛顿语),所以牛顿称 $\frac{\Delta x}{\Delta t}$ 为最初比,然后让 $\Delta t$ 逐渐趋近 0(请注意,牛顿在这里不但使用了趋近的说法,还多次用了"极限"一词,只不过那时并没有明确的极限概念,牛顿也没有花功夫去说明极限的含意是什么),于是最初比有一个"极

限"$\dot{x}(t)$，牛顿称为"最终比"，牛顿把变化中的量如 $x(t)$ 称为流量(fluent)，而把其变化率 $\dot{x}(t)$ 称为流数(fluxion).牛顿的方法论似乎有些混乱，有时他也使用不可分量，但是例如在《原理》一书中他又批评不可分量，明确地说："如果我说某量由粒子组成，……不要以为我是指不可分量，而是指趋于0的可分量"，牛顿以为用运动学的观点去解释这些问题就可以避免古希腊人的思辨方法、形而上学方法的困难.他做到了吗？下面再引牛顿的几段话(这些话都见于《原理》中文版38页.)："可能有人反对，认为不存在将趋于0的量的最后比值，因为在量消失之前，比率就不是最后的，而当它们消失时，比率也没有了"(就是说，$\Delta t\neq 0$ 时就不是瞬时速度，$\Delta t=0$ 就连运动也没有了)."最后速度……就是……物体到达其最后处所并终止运动时的速度……最后比可以理解为……它消失的那一瞬间的比……在这运动尚存的最后时刻速度有一极限，不能超越，这就是最后速度，"牛顿还说"我论及最小的、将消失的或最后的量，读者不要以为是在指确定大的量，而是指作无止境减小的量".所谓瞬间是什么？"它消失那一瞬间"是什么？什么叫"无止境减小的量"？是不是0？可不可以略去不计？这才是问题所在.

看一下切线问题还会更清楚，下面我们讲一下费马的作法，图1-5画出了求 $y=f(x)$ 之图像在 $P$ 点的切线的方法，此图含意自明，下面除了其中关键的 $PR$ 记为 $E$ 之外均不再解释.令 $G$(即 $x$ 点)得一增量 $E$(就是我们常用的 $\Delta x$)而变为 $G_1$，由于$\triangle TGP\backsim\triangle PRT_1$，所以

$$TG:PG=GG_1:T_1R=E:T_1R.$$

但是，当 $E$ 非常小时，费马说 $T_1R$ 和 $P_1R$ 相差不多，从而可以把 $T_1P_1$ 略去不计，用我们熟悉的语言就是当 $\Delta x$ 很小时，$\mathrm{d}y$ 与 $\Delta y$ 之差可以忽略不计，所以

$$\begin{aligned}TG:PG&=E:(P_1G_1-PG)\\&=E:[f(x+E)-f(x)]\end{aligned}$$

但 $PG=f(x)$，故上式给出

$$\begin{aligned}TG&=Ef(x)/[f(x+E)-f(x)]\\&=f(x)\Big/\frac{f(x+E)-f(x)}{E}.\end{aligned}\tag{7}$$



图1-5

最后费马说，"去掉 $E$ 项"(即令 $\Delta x=0$)，立即可得 $TG$，在确定了 $T$ 点位置以后，连接 $TP$ 即得切线.

这里遇到的问题是：令 $E=0$ 时，(7)之分母成了$\frac{0}{0}$而没有意义.然而要注意，在那个时代，人们知道的函数不多，例如笛卡儿就只讨论 $f(x)$ 是多项式的情况.这时 $f(x+E)-f(x)$ 仍是多项式，而且每一项都有因子 $E$，因此，(7)之分母中 $E$ 可以约去(注意，这当 $E=0$ 时是不合法的)，因而不会出现$\frac{0}{0}$问题了.但是令 $E=0$ 不但是做了不合法的事，而且 $E=0$ 就没戏可唱了，这和上面讲的 $\Delta t=0$ 就没有运动如出一辙.$E$ 又是0又不是0，是一个方生未死似0而又非0的无穷小量！可是什么是无穷小呢？再说上面略去了 $T_1P_1$，究竟什么样的东西可以略去不计呢？所以，情况仍与芝诺时代差不多，必须要回答什么是无穷小，而且不能只用思辨的方法，要解决具

体问题，如当年用穷竭法计算圆面积一样.

我们不能像卡瓦列里说的那样，把严格性完全交给哲学家.《几何原本》对这个时代的数学家仍然是不可改变的规范.例如牛顿的基本著作《自然哲学的数学原理》就完全是模仿《几何原本》的.牛顿知道其中的问题时常出在无穷小量上面：说它不是零，并用看起来很像零的英文小写字母 $o$ 来记它，又常用 $a+o=a$，或在两个无穷小乘积中令它为零.说它是零，又允许以它为分母，并且称$\frac{0}{0}$为不定式.18 世纪的许多卓越的数学家都为此绞尽了脑汁.这当然瞒不过哲学家的法眼.这里必须提到爱尔兰克罗因地方的主教伯克莱(Bishop George Berkeley，1685—1753)的一本著名的小册子《分析学家，或致一位不信神的数学家》(The Analyst，or a Discourse Addressed to an Infidel Mathematician)，读者可以在前面引用的李文林主编的《数学珍宝》一书 316～321 页上看到此文的摘译，下面引用的译文均来自此书."不信神的数学家"是指哈雷(E. Halley)，哈雷彗星的发现者，对牛顿写出《原理》一书起了重要的作用，并资助出版.伯克莱写这本书当然是为了宣扬自己的宗教观和自己的哲学.大体说来，他的论据就是：连牛顿的微积分、无穷小量那样模糊不清，逻辑混乱的东西都可以相信，为什么你们却不肯相信上帝呢？我们不想讨论哲学，但是应该指出，伯克莱的批评是切中要害的.伯克莱直接指名责备牛顿"曲线求积术"一文中给 $x$ 以非零的增量$h$，然后作了一些代数运算，然后又叫 $h=0$，这岂非背离了逻辑学中最起码的矛盾律(即 $A$ 不是非$A$)吗？而在神学中是不允许这样做的.而且他两次引用牛顿在该文中的原话："在数学中最微小的误差也不可以忽略"来讥讽牛顿.伯克莱指出，牛顿的"流数"(即我们今天讲的导数)是不可理解的，更何况二阶、三阶……流数"全都超出整个人类理解能力".这种又是零又不是零的增量，还有增量的增量……"这就是我们的现代数学家们……全部思想的基石".可是这些全是不可理解的东西.至于由此得到的成功，伯克莱说："他使用流数，就像建筑中使用脚手架一样，"房子盖好了自然"就立即将它们弃之一边.""我所非议的不是您的结论，而是您的逻辑和方法：您是怎样进行证明的？您所熟知的对象是什么？关于它们您的表述是否清楚？您依据的原理是什么？它们是否可靠？您是如何应用它们的？"(着重点是本书作者加的——作者)至于得到了正确的结论，这是"因为这个错误被另一个相反的但程度相当的错误抵消了."(这句话是指莱布尼茨.)伯克莱正确地点出了要害：那个似零又不是零的 $h$ 是什么？"这些消逝的增量又是什么呢？它们既不是有限量，也不是无限量，又不是零，难道我们不能称它们为消逝量的鬼魂吗?"其实，从我们上面引述的牛顿在《原理》一书中对自己的最终比的说明，并且一再强调自己的方法与不可分量并不相同.看来，难道不能认为牛顿自己也感到了类似的问题，因而理不直，气不壮，处于彷徨与无奈之中吗?

把微积分放在一个严格的基础上是很长一段时间数学家着重努力的大问题.但是有一点应该指出，这一段时期所有的数学家是同时用极大的力量从事扩大与深化微积分的应用的.现在的大学数学系，一直到高年级的许多课程的内容都是在 18 世纪、19 世纪形成的，这一点与希腊时代非常不一样.这是生产力迅猛发展及其带来的科学的迅猛发展的表现.达朗贝尔有一句名言，很可以反映当时的情况："前进吧，你会得到信心!"

确实是这样，数学进入了一个高歌猛进的时代.新思想，新方法层出不穷，新领域一天天被发现、被征服.然而却全处在伯克莱大主教提出的一长串问题的笼罩下.我们想从欧拉的《无穷小分

析引论》①中转述一大段,而且仍然沿用欧拉的记号.这一段正是讲的指数函数与对数.

任取一数 $a$ 为底,则 $a^0=1$.若 $\omega$ 表示一个无穷小,则 $a^\omega=1+\Psi$,$\Psi$ 也是一个无穷小.所以我们可设 $\Psi=k\omega$,而

$$\omega=\log_a(1+k\omega).\tag{8}$$

例如当 $a=10$ 时,$k=2.302\ 58$.对任意数 $j$ 有

$$a^{j\omega}=(1+k\omega)^j=1+\frac{j}{1}k\omega+\frac{j(j-1)}{1\cdot 2}k^2\omega^2+\cdots.$$

令 $j=\dfrac{z}{\omega}$ 而 $z$ 为定数,则 $j$ 等于无穷大,代入上式有

$$a^z=\left(1+\frac{kz}{j}\right)^j=1+\frac{1}{1}kz+\frac{j(j-1)}{1\cdot 2}\frac{k^2z^2}{j^2}+\frac{j(j-1)(j-2)}{1\cdot 2\cdot 3}\left(\frac{kz}{j}\right)^3+\cdots.\tag{9}$$

但因 $j$ 是无穷大,所以 $\dfrac{j-1}{j}=1$,$\dfrac{j-2}{j}=1$ 等等,代入上式,

$$a^z=1+\frac{kz}{1}+\frac{k^2z^2}{1\cdot 2}+\cdots,\tag{10}$$

令 $z=1$ 有

$$a=1+\frac{k}{1}+\frac{k^2}{1\cdot 2}+\cdots.\tag{11}$$

上面说到 $\Psi=k\omega$,取 $a$ 为相应于 $k=1$ 的那个数,则由(11)式有

$$a=2.718\ 281\ 828\ 459\ 045\ 235\ 360\ 28.\tag{12}$$

(请注意,欧拉时代没有计算机,他怎么能用手算出 e 到 23 位小数?)我们记这个数为 e.下面一段是欧拉的原话:

> "为简单计,我们用符号 e 表示此数:$e=2.718\ 281\ 828\ 459\cdots$.它是自然对数或称双曲对数的底,它相应于 $k=1$,而且 e 表示无穷级数 $1+\dfrac{1}{1}+\dfrac{1}{1\cdot2}+\dfrac{1}{1\cdot2\cdot3}+\dfrac{1}{1\cdot2\cdot3\cdot4}+\cdots$ 之和".

再看对数,由(9),令其中的 $a=e$,从而 $k=1$,有

$$e^z=\left(1+\frac{z}{j}\right)^j=1+\frac{z}{1}+\frac{z^2}{1\cdot 2}\cdot\frac{j}{j}\cdot\frac{j-1}{j}+\cdots.\tag{13}$$

如果把 $e^z$ 写成 $1+x$,则由上式

$$\begin{aligned}z&=j\left[(1+x)^{\frac{1}{j}}-1\right]\\&=j\left[\frac{x}{j}+\frac{x^2}{1\cdot 2}\cdot\frac{1}{j}\left(\frac{1}{j}-1\right)+\frac{x^3}{1\cdot 2\cdot 3}\cdot\frac{1}{j}\left(\frac{1}{j}-1\right)\left(\frac{1}{j}-2\right)+\cdots\right]\\&=x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots.\end{aligned}$$

但是 $z$ 就是 $\log_e e^z=\log_e(1+x)$,所以又有

① 欧拉这一著作,是用拉丁文写的,书名 Introductio In Analysis Infinitorum,1748.应译为《无限量分析引论》,1988 年,J. D. Blantan 有英译本,就采用了这样的书名:Introduction to Analysis of the Infinite,我们仍然从俗把欧拉这部著作译为《无穷小分析引论》

$$\log_e(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots. \tag{14}$$

我们能同意伯克莱大主教说的:"所有这些错误碰巧都抵消了,所以欧拉得出了如此惊天动地的结果"吗? 当然不能,欧拉是一位极为罕见的大数学家,他在处理数学与力学问题上的能力是非凡的.如果数学中以欧拉命名的东西都没有了,数学史一定要改写.但是数学仍然会发展,而是一定会和今天一样好.欧拉的成就应该归功于他非凡的洞察力和数学运算能力.更深一层来看,牛顿、莱布尼茨真正伟大之处在于他们真正抓住了变量的数学的本质,形成了高屋建瓴势如破竹的局面.所以在一两百年之内巨人辈出也正是时势造英雄.余下的工作就是把问题都搞清楚,否则数学是不可能再进步了.请读者看欧拉的一段文章目的就是请大家注意,欧拉没有一个结论是错的,但是真把他的毛病都消除,却是要大费周章.特别是由二项级数,再逐项求极限决非轻而易举.读者们读过的微积分教本上面那一小段引文要花多少篇幅才说得清楚,读者自己都有体会.

上面我们完全没有讲莱布尼茨的贡献.这不是由于在"发明"微积分的"知识产权"问题上我们有什么偏爱.而是由于,一开始我们就说了,本章并不是一个初步但比较全面的微积分历史.莱布尼茨和牛顿虽然在微积分的理论和方法上都不一样,但是在什么是微分 $dx$,$dy$(这都是莱布尼茨的记号)与不可分量、无穷小量的关系上,他与牛顿遇到相同的困难.伯克莱大主教在反对他们二人上似乎也是一视同仁的,所以我们就完全略去这一部分.好在在本章开始时介绍给读者的两本书中都有了介绍.

解决遗留下来的问题——其核心从希腊时代起就是关于无穷小量与不可分量等等的问题——出路在于极限理论.极限的概念牛顿等许多人都已经有了,而在柯西以前,最接近当代极限概念的是达朗贝尔.然而所有这些论述都过于几何化:所谓一个变量 $x$ 趋于 $a$,人们自觉或不自觉地是从运动学的角度来领会的.因此柯西最大的功绩在于把极限的研究算术化了.这方面的先驱应该指出的还有波尔察诺(Bernhard Bolzano, 1781—1848)和阿贝尔(Niels Henrik Abel, 1802—1829)前者曾指出高斯关于代数的基本定理(每个代数方程都有根存在)证明不够严格,因为高斯利用了连续函数的中间值定理未加证明.他又是第一个给出不可求导的连续函数的例子(1848)的数学家,比魏尔斯特拉斯早 30 年.他多年来是布拉格的一位教士而未为人知.阿贝尔很年轻就去世了,一生贫病交加.他对当时分析中缺少严格性可说是痛心疾首.真正成就和影响最大的当然是柯西,他写了几本教本:《工科大学分析教程》(Cours d' Analyse de l' École Politechnique,1821).《无穷小计算概要》(Résumé des Leçons sur les Calculs Infinitesimals, 1823)与《微分学讲义》(Leçons sur les Calculs Differentiels,1829),使我们看到了今天我们熟知的微积分教本.柯西所作的事就是使极限摆脱几何的束缚.极限概念如果不能摆脱几何的束缚,终将会引起混乱.例如说圆内接正 $n$ 边形当 $n\to\infty$ 时极限就是圆,那么就可以问,是否终于变成了圆? 圆的性质全都含在正多边形的性质之中了? 柯西明确指出,极限就是一个数,所谓变量 $x$ 趋向 $a$ 时 $f(x)$ 以 $A$ 为极限就是:若代表变量 $x$ 的一串数值无限地趋近固定数 $a$ 且其差可以任意地小时,$f(x)$ 与 $A$ 之差也可以任意小.但是稍后的数学家如魏尔斯特拉斯仍然对此定义不满,认为仍然包含了几何学与运动学的痕迹,应该更进一步算术化.变量是什么? 魏尔斯特拉斯给的回答就是:若有一个**无限**非空集合 $D$(例如实数集),而 $x$ 可以取 $D$ 中任一元为值,$x$ 就称为一个变量.$\lim\limits_{x\to a} f(x) = A$ 的意思就是:任给 $\varepsilon>0$ 必可找到 $\delta=\delta(\varepsilon)>0$,使当 $|x-a|<\delta$ 时,$|f(x)-A|<\varepsilon$.我们

看到，这样做其实正是发展了《几何原本》中把穷竭法归结为不等式的实践，于是许多问题自然化解了：这些正 $n$ 边形最后会不会与圆重合？根本不必问这样的问题，反正有 $\lim_{n\to\infty} P_n = S$. 每一个 $P_n$ 均不是 $S$，$\{P_n\}$ 中没有最后一个元."最终比"是不是运动停止的那一瞬间的速度？无所谓这一瞬间那一瞬间，既无所谓运动开始的最初比，也无所谓运动停止"那一瞬间"的最终比.我们看到的只有 $\frac{\Delta x}{\Delta t}$ 是 $\Delta t$ 的函数，$\Delta t$ 与 $\frac{\Delta x}{\Delta t}$ 都是变量，而当 $\Delta t\to 0$ 时 $\frac{\Delta x}{\Delta t}$ 也有极限 $\lim_{\Delta t\to 0}\frac{\Delta x}{\Delta t}$，它就叫做瞬时速度.运动可以还继续，无所谓最终.什么是无穷小？它没有任何特殊的神秘之处，用不着与"单元"、"原子"、"不可分量"……扯到一起，它只是以 0 为极限的变量而已.

先是柯西，再继之以魏尔斯特拉斯，把整个微分积分放在极限的基础上，而后来又放在不等式的基础上，可以说，魏尔斯特拉斯如愿以偿地"清洗"了运动学的"痕迹"，变量的数学静态化了(既已静态化，也就用不着再为芝诺的悖论操心了).这就是读者们学过了的微积分.

于是我们看到，什么是导数？是极限；什么是积分？也是极限；什么是 $\sum a_n$？还是极限.(只有什么是微分？这比较麻烦，第三章全章就是为了回答这个问题.)什么是极限？也很清楚，极限是数.那么什么是数？这就追溯到古希腊的第一次危机.请看康托尔(G. Cantor)的回答："$\sqrt{3}$ 只是一个尚未确定的数的符号，而不是它的定义.然而，用我的方法可以满意地给出其定义为

$$\{1.7, 1.73, 1.732, \cdots\}".$$

康托尔的话意思是，实数需要定义.其方法之一是定义实数为一个具有特定性质的有理数序列.这样一来，单位正方形的对角线长也是一个"数"，其定义是一个序列：

$$\{1.4, 1.41, 1.414, \cdots\}.$$

用这样的方法解决了毕达哥拉斯的困惑，使得直线上每一个点都对应一个数，每一个数也对应一个点，即使得点与点的坐标一一对应.这是解析几何的基础，而且把整个几何"算术化"了.例如两个线段的比现在很容易定义了，若 $l_1$，$l_2$ 的长是有理数，则 $l_1:l_2=\frac{l_1}{l_2}$ 也是有理数；若 $l_1$，$l_2$ 之长有一个是无理数，$l_1:l_2=\frac{l_1}{l_2}$ 仍有意义，无非可能是无理数而已.总之，量与数的对立消除了.不仅如此，有关极限理论最困难的定理也都得到了证明.这就是魏尔斯特拉斯所企盼的把整个微积分算术化.我们将在第六章详细讨论这些问题.

总之，微积分现在已经建立在稳固的基础上了.它有如一个"平台"，在这个平台上可以建立许多大厦.可以用它来解决许多以前所不能设想的问题.我们这本书就想对这个平台上的一些"大建筑"作一番漫游.当然，数学科学还有许多其它的平台，即具有自己的新思想、新方法的领域.我们虽然不一定能去参观，不一定能去"定居"，但是可以告诉读者，这些平台中大多数与牛顿的平台有密切的关系.而且从数学的历史来看，如果想找一个比牛顿平台更基本的，可能就只有《几何原本》了.

不过还有一个问题要说几句，既然微积分已建立在可靠的基础上了，是否由此就完全否定了直觉的作用？我们不打算进入哲学的领域.从数学的教学和研究来看，我们还只能说，这么多世纪的数学发展证明了人们的直觉还是可靠的，需要的只是深入的分析.直觉仍是数学的思想与理论的来源之一.例如芝诺的悖论，我们还只能说，如果不是终日冥想，而是进入具体的数学与物理学的领域，人们就不会只作这种抽象的玄想.从芝诺悖论会产生集合论的问题，也都在集合论作

为数学的一个分支的框架内，或者得到解决，或者得到深化．又如无穷小，运动等等，我们只要知道了在数学中是如何处理这些问题的，就会发现这些直觉不但无害而且有益．而且，数学懂得越多，从直觉里得到益处也越多．至少是受“严格性不足”之害越少．所以直到现在，人们还是说“设某个增量 $\Delta x$ 是无穷小”，说什么某个小体积是“无穷小元”，或者“微元”之类的话．本书也是这样做的．总之，越往后走，越深入数学和物理学，您会得到越大的“自由”，而不必谨小慎微．

大约二百年前，达朗贝尔用一句鼓舞人的口号来描述当时的数学：“前进吧，你就会有信心”．而现在的情况则是：人类正充满信心，人类正在前进！

# 第二章　函　　数

微积分研究的对象主要是函数. 其主要的动机并不是思辨的需要，而时常是更实际的经济利益的驱动. 例如三角学，自古希腊就有系统的研究. 这是为了研究天体的运行，也是为了航海的需要. 所以当时不只是有平面三角学还有球面三角学. 随着资本主义时代的来临，人类活动的广度与深度大为增加，例如对数的出现显然是为了简化航海中的计算. 值得注意的是，这些活动甚至直接影响到学术界的最上层. 下一章开始时我们会讲到牛顿，还有哈雷对天体运动的研究与航海的关系. 现在我们将指出，对数的一位早期创始者尼古拉•麦卡脱[1]实质上是用积分 $\int_0^x \frac{\mathrm{d}x}{1+x}$ 来定义对数. 牛顿接受了他的思想，并把自己最得意的二项级数用上，与麦卡托互相独立地得到了 $\log(1+x)$ 的泰勒级数展开式. 然后又用非常精巧的反演方法，得到了 $\mathrm{e}^x$ 的幂级数展开式. 牛顿也用同样的方法来研究 $\arcsin x=\int_0^x \frac{\mathrm{d}x}{\sqrt{1-x^2}}$ 以及它的反函数 $\sin x$. 一直到 1715，泰勒(Robert Taylor，1685—1731)发表了“增量法及其逆”一文，才给出了我们现在知道的泰勒公式. 所以，历史的发展与我们现在的教科书有些矛盾，先有反函数 $\log(1+x)=\int_0^x \frac{\mathrm{d}x}{1+x}$ 和 $\arcsin x=\int_0^x \frac{\mathrm{d}x}{\sqrt{1-x^2}}$ 的级数展开式，后来才有 $\mathrm{e}^x$ 和 $\sin x$ 的级数展开式. 为什么现在通用的微积分教本要用与历史不符的讲法呢？一个重要原因是教学的需要. 首先看到这一点的欧拉，把关于微积分基础的讲授放在前面作为预备知识，于是多年以后三角学完全在中学里了. 大学里先是给出了 $\mathrm{e}=\lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n$，作为有界上升数列必有极限的一个重要的例子. 可是为什么重要？为什么以 e 为底的对数称为“自然”？都只好不讲了. $\mathrm{e}^{\mathrm{i}\theta}=\cos\theta+\mathrm{i}\sin\theta$ 按逻辑的要求应该放到另一门课去讲，因为必须先定义什么是复变量的幂级数(其实这只是一句话的问题). 现在，我们既然假设读者都学过微积分，就有可能用一种比较接近历史真实的、比较接近物理学的讲法. 我们的出发点是：一个函数 $y=y(x)$ 可以作为一个微分方程的柯西问题

$$\begin{cases}\dfrac{\mathrm{d}y}{\mathrm{d}x}=f(x,y),\\ y(x_0)=y_0\end{cases}\tag{A}$$

之解来定义. 这里 $y$ 可以是一个函数，也可以是一个向量 $y=(y_1,\cdots,y_n)$(当然 $f$ 也相应地应为

---

[1] Nicholas Mercator，1620—1687，数学史上有两位麦卡脱，另一位是 Gerhard Kremer Mercator，1512—1594 是著名的地图制作家，现在地图上使用的麦卡脱投影法就是他创造的. 他为制作地图所引进的一些名词，现在我们仍然使用在微分流形理论中，见第七章. 这二位的姓名中 Mercator 一字在德文文献中常作 Kaufman，商人之意. 其实 Mercator 就是商人一词的拉丁文说法，英文的 Merchant 也来自此词.

一个向量 $f=(f_1,\cdots,f_n)$). 其实欧拉以及现在许多深一点的微积分教本都是这样作的，特别当涉及复变量时是如此. 那么它是否能定义一个(而不是许多个)函数(也不是什么都没有定义)呢？从数学上讲这就是常微分方程的柯西问题的存在与唯一性定理. 为此下面的条件就足够了：

**常微分方程的柯西问题(A)的存在与唯一性定理** 设(1)右方的向量值函数 $f(x,y)$ 的每一个分量均定义在 $\Omega$ 中，$(x_0,y_0)\in\Omega$ 而且在 $\Omega$ 的子集 $|x-x_0|\leqslant a,|y-y_0|\leqslant b$ 中对 $(x,y)$ 连续对 $y$ 满足利普希茨条件，则(A)必有唯一解存在于 $[x_0-h,\ x_0+h]$ 中. 这里 $h=\min\{a,b/M\}$，$M=\sup\limits_{|x-x_0|\leqslant a,|y-y_0|\leqslant b}|f(x,y)|$.

这个定理我们就不来证明了. 不过希望读者注意，这是一个很细致的定理，在实际用起来的时候一些似乎是很不起眼的地方，例如 $h$ 的大小，都会出问题.

我们上面已经说了欧拉就这样来研究过一些函数. 但是欧拉的时代并没有这样的定理. 这倒不是因为这个定理的证明很难，而是因为当时人们并不认为需要证明这样的定理，而且证明这个定理时必不可少的收敛性的概念，欧拉时代的数学家并不明白. 例如看一个质点的运动，按牛顿第二定律，运动方程就是一个常微分方程组，如果给出了一定的初始条件，则该质点的运动状况就完全确定了. 而且欧拉的想法是：把时间分成一小段一小段. 在第一小段(从 $x=x_0$ 到 $x=x_0+h$)，设 $f(x,y)=f(x_0,y_0)$，于是得到 $y=y_0+f(x_0,y_0)(x-x_0)$，这样到 $x=x_1=x_0+h$ 时，$y=y_1=y_0+f(x_0,y_0)h$. 在第二小段时间，从 $x=x_1$ 到 $x=x_1+h$ 中不妨用 $f(x_1,y_1)$ 代替 $f(x,y)$，而得 $y=y_1+f(x_1,y_1)(x-x_1)$. 仿此进行下去，我们会得到一条折线，称为欧拉折线. 欧拉时代的人们认为 $h\to 0$ 时就会得到真实的运动状况，而没有想到收敛性是要证明的. 所以到18世纪为止，人们实际上是以物理的分析代替数学的分析. 而且从哲学上看，这样的想法没有任何问题，因为世界上的一切都服从某种决定论，存在与唯一性定理无非是哲学上的决定论的数学表示. 哲学上对了，数学上也就犯不着再自寻烦恼了. 顺便说一下，不但欧拉折线是欧拉的创造，连牛顿的运动定律可以写成微分方程(A)也始自欧拉. 所以我们常讲的牛顿运动方程也不妨称为牛顿—欧拉方程. 下一章§6我们还要分析这里涉及的方法论问题，但是到了19世纪就发现这样做在数学上会出毛病. 首先提出要证明存在与唯一性的是柯西. 他在1820—1830年间发现仅有 $f(x,y)$ 的连续性并不足以保证唯一性，于是他补充上“$f_y(x,y)$ 也连续”的条件. 利普希茨(Lipschitz，1832—1903)在1876年把这个条件减弱为我们现在熟知的利普希茨条件. 人们还找到了不符合这类条件就会破坏唯一性的反例. 而为了证明存在性，则 $f(x,y)$ 的连续性就够用了.

我们的读者都受到过相当程度的逻辑推理的训练，当然会问：这样做会不会是循环论证？肯定不会. 在证明上述定理时我们没有用到任何的具体函数——不论是指数对数还是三角函数、反三角函数. 还会问：这样做岂非杀鸡用牛刀？我们重新回到历史走过的老路，是为了看一下，微积分的发展怎样帮助我们更深刻地认识大自然的规律. 从这个意义上讲，指数函数当然不是“鸡”，说它是“龙”也不为过. 总之这不只是特例，而是十分重要的数学模型.

## §1 增长的数学模型：指数与对数

**1. 指数函数及其基本性质** 在自然界和人类社会中有许多量的增长之快慢与该量现在的

值成正比.例如人在银行的存款,在利率一定的条件下,存款增长多少,亦即利息之多少是与现在存款量(作为本金)成正比的.这里存款计息假设是按一定时期(年、半年)来计算的.在计息期末计息之前,存款额我们约定按期初之值计算,即是本金.又如某种细菌,某一个菌株是否恰好在一个时刻分裂可以是随机的.但在一定条件下,在一段时间中细菌群体的一个百分比(设为 $\lambda$)总是在分裂的,细菌就这样繁殖.某一定量的放射性物质的蜕变也是这样.对这种现象,既可以把它作为连续变化过程来处理(例如细菌的分裂,放射性元素原子的蜕变),也可以把它作为离散现象来处理.例如银行存款的增长是离散的,即使存期不到,也是按天计算,不会对时间连续计算.但是下面我们会看到,即使这个问题也可以作为连续变化的量来看待.总之,对这类问题可以从连续变量与离散变量两个角度来讨论.下面我们首先从连续变量的角度来讨论.

从连续变化的角度看,上述问题的数学模型是

$$\begin{cases}\dfrac{\mathrm{d}y}{\mathrm{d}t}=\lambda y, & (1)\\ y(0)=y_0. & (2)\end{cases}$$

这个问题可以化简到 $\lambda=1$(因为只需作一个变换,即引入新的自变量 $t_1=\lambda t$ 即可)的情况,又可以假设 $y_0=1$.因为若得到了 $y_0=1$ 的解 $E(t)$,则 $y_0E(t)$就是问题(1),(2)之解.从上面讲到的存在与唯一性定理立即可知这时(1),(2)有解 $E(t)$存在,而且 $E(t)$在其定义域上是连续可微函数.从我们熟知的常微分方程知识立即可知

$$E(t)=\mathrm{e}^t, \tag{3}$$

这里 $\mathrm{e}=\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n$,但因我们目前还没有定义 e,所以暂时还不知道(3)式是什么意思.但是我们有极为重要的

**定理 1**(加法定理)　上述 $E(t)$之定义域是 $\mathbf{R}$: $-\infty<t<+\infty$,且在其定义域内有

$$E(t_1+t_2)=E(t_1)E(t_2),\ t_1,t_2\in\mathbf{R}. \tag{4}$$

**证**　用读者已熟悉的初等方法自然可得以上一切结论,但是我们现在不能用(3)式,而要回到微分方程(1).如果用初等方法求积(1)(令 $\lambda=1$)

$$\frac{\mathrm{d}y}{y}=\mathrm{d}t,$$

双方对 $t$ 求积分,就应考虑到如果 $y=y(t)$在某个 $t$ 值处为 0 怎么办?所以严格的证法应该分别两种情况:一是 $y(t)$处处不为 0,这时可用求积法来处理它.但要证明相应于这个 $y(t)$的 $t$ 值充满 $\mathbf{R}$ 并非易事.二是 $y(t)$在某处为 0.所以,初等的求积方法其实不是一个好方法.我们还可以应用读者可能已学过的一般的存在唯一性定理.在那里,我们证明了解在$[x_0-h,x_0+h]$中存在,而 $h=\min\left(a,\dfrac{b}{M}\right)$.但是这里也有困难.因为 $a,b$ 固然可以取得任意大,$M$ 也会随之增大.所以在一般的常微分方程解的存在与唯一性定理后,专门有结果指出,对于线性方程(我们用矩阵形式来表示)

$$\frac{\mathrm{d}y}{\mathrm{d}t}=A(t)y,$$

若 $A(t)$在区间$[a,b]$上连续,则解的定义域必包含$[a,b]$,所以对于常系数线性方程,解的定义域必为 $\mathbf{R}$.这一切都属于比较细微的地方而可能会注意不到.我们下面要利用常系数线性方程的

平移不变性来证明它.加法定理本身也就是平移不变性的表现.

设 $E(t)$ 已定义在 $[t_0-h, t_0+h]$ 中，$t_0=0$，于是令 $t_1=t_0+h_1, h_1=h-\varepsilon$，$\varepsilon$ 是任意小正数，并记 $E(t_1)=E_1$，于是 $E(t)$ 在 $[t_0, t_0+h]$ 上也是柯西问题

$$\frac{dy}{dt}=y,\ y(t_1)=E_1$$

之解.令 $t=\tau+t_1$，则此问题又可化为

$$\frac{dy}{d\tau}=y,\ y\Big|_{\tau=0}=E_1.$$

由前所述，其解为 $E_1E(\tau)$，而因为新的柯西问题形状与 (1)，(2) 相同，故由解的唯一性定理，在 $[t_1, t_1+h]$ 上，$E(t)=E_1E(\tau)=E(t_1)E(t-t_1)$ 与 (1)，(2) 完全相同，所以这个解应可以定义于 $\tau\in[0,h]$ 上，即是说 $E(t)$ 不仅可以定义在 $[t_0, t_0+h]$ 上，而且可以定义在 $[t_1, t_1+h]=[t_0+h_1, t_0+h+h_1]$ 上，再由 $\varepsilon$ 之任意性，知 $E(t)$ 可以定义到 $[t_0+h, t_0+2h]$ 上.仿此进行下去知 $E(t)$ 可以定义到 $[t_0, +\infty)$ 上，同理也可定义到 $(-\infty, t_0]$ 上.总之 $E(t)$ 之定义域为 $(-\infty, +\infty)$.

由前面已证的

$$E(t)=E_1E(\tau)=E(t_1)E(t-t_1),$$

令 $t=t_1+t_2$，由于 $E(t)$ 之定义域已知是整个 $\mathbf{R}$，所以 $t_1, t_2$ 取任意实数值均是允许的.这样，上式就给出了加法定理：对任意实数 $t_1, t_2$ 均有

$$E(t_1+t_2)=E(t_1)E(t_2).$$

定理证毕.

上面证明的关键在于引入了新的自变量 $\tau$ 以后，柯西问题的形状不变，因而原来的定理现在仍然适用.这就是平移不变性.如果 (1) 表示某种规律，则当 $t$ 表示时间时，平移不变性就表示这个规律在任何时间都相同；如果 $t$ 表示空间位置，平移不变性就表示这个规律在各地都一样，常系数的微分方程就表示这类规律，因此是最重要的.

下面要问，$E(t)$ 究竟是什么函数？准确一些说，应该问，$E(t)$ 与我们过去熟知的那些函数有何关系？下面要证明 $E(t)$ 是指数函数.

1. 首先注意到 $E(t)$ 恒不为 0，因为若 $E(t_0)=0$，不妨设 $t_0=0$（令 $t=\tau+t_0$，再用平移不变性），则有 $\frac{dE(\tau)}{d\tau}=E(\tau)$，$E(0)=0$，由解的唯一性定理应有 $E\equiv 0$，但因 $E(0)=1$，所以 $E(t_0)=0$ 不能成立，从而 $E(t)$ 恒不为 0. $E(t)$ 既是连续函数，故知 $E(t)>0$.

2. $E(t)$ 既然定义在 $\mathbf{R}$ 上，则它把 $\mathbf{R}$ 映到什么地方？因为 $\frac{dE}{dt}=E>0$，所以 $E$ 是上升函数，它必把 $\mathbf{R}$ 单调地映到某区间 $(E(-\infty), E(+\infty))$ 上，所以要问 $E(\pm\infty)=?$ 由于

$$\int_1^E \frac{dE}{E}=\int_0^t dt,$$

令右方的 $t\to+\infty$，则得 $\int_1^E \frac{dE}{E}$ 在 $t\to+\infty$ 时趋于 $+\infty$，为此需要 $E\to+\infty$，所以 $E(+\infty)=+\infty$. 当 $t\to-\infty$ 时，因为 $\int_1^E \frac{dE}{E}$ 在 $E=0$ 处发散，易见这时 $E\to 0$，即是说 $E(-\infty)=0$. 所以 $E(t)$

把 $\mathbf{R}$ 单调地映为 $(0, +\infty)$. 下面讨论 $E(t)$ 的反函数时，这个性质是其基础.

3. 最后证明 $E(t)=[E(1)]^t$. 首先设 $t$ 是正整数，例如 $t=2$.

$$E(2) = E(1+1) = E(1)\cdot E(1) = E(1)^2.$$

同理 $E(n)=[E(1)]^n$，$n$ 为正整数.

其次看 $E(-n)$，$n$ 仍为正整数. 注意到由定义：$E(0)=1$（就是 (2) 式），所以

$$1 = E(0) = E(n+(-n)) = E(n)\cdot E(-n) = [E(1)]^n E(-n),$$

而 $E(-n)=[E(1)]^{-n}$. (注意前已证明 $E(t)$ 恒不为 0，故 $E(1)\neq 0$，这样，上面的推导才有意义.) 总之

$$E(n) = [E(1)]^n, n = 0, \pm 1, \pm 2, \cdots.$$

再看 $t=\frac{n}{m}$ 为有理数 ($m$，$n$ 为整数，设 $m>0$) 的情况. 首先，由 $1=\frac{1}{m}+\cdots+\frac{1}{m}$ ($m$ 项)，有

$$E(1) = E\left(\frac{1}{m}+\cdots+\frac{1}{m}\right) = \left[E\left(\frac{1}{m}\right)\right]^m,$$

$$E\left(\frac{1}{m}\right) = [E(1)]^{\frac{1}{m}}.$$

于是对 $n>0$，$\frac{n}{m}=\frac{1}{m}+\cdots+\frac{1}{m}$ ($n$ 项)，有

$$E\left(\frac{n}{m}\right) = \left[E\left(\frac{1}{m}\right)\right]^n = [E(1)]^{\frac{n}{m}}.$$

而 $n<0$ 时，则利用 $E\left(\frac{n}{m}+\frac{(-n)}{m}\right)=E(0)=1$，从而 $E\left(\frac{n}{m}\right)=\left[E\left(\frac{(-n)}{m}\right)\right]^{-1}=\left[E\left(\frac{1}{m}\right)\right]^{(-n)(-1)}=[E(1)]^{\frac{n}{m}}$ (注意 $-n>0$)，所以 $t$ 为有理数时 $E(t)=[E(1)]^t$ 也成立.

最重要的一步是 $t$ 为一般实数的情况. 在通常的微积分教本中定义 $a^t$ 时比较马虎一些的就说 $a^t$ 是连续函数. (对这句话则不加论证，而且 $a^t$ 尚未定义又怎么知道 $a^t$ 是连续函数呢？) 在比较严肃一些的教本中都是取一串有理数 $t_n\to t$，然后讨论 $\lim\limits_{n\to\infty}a^{t_n}$，当 $a\neq 1$ (现在不会出这样的事，因为 $E(t)$ 是上升函数而于 $t>0$ 时，$E(t)>E(0)=1$) 例如可以取一串上升的 $t_n\to t$，再用单调有界数列定理证明这个极限存在. 同时还要证明，不论 $\{t_n\}$ 如何取，不论它是否单调，$\{a^{t_n}\}$ 都有相同的极限. 用这个极限值作为 $a^t$ 之定义，这样才把指数函数 $a^t$ 定义在整个 $\mathbf{R}$ 上. 总之是用上了实数理论的全武行！本书第六章会详细讲实数理论，而且为了以后考虑一般函数空间，是用柯西序列来定义实数的. 但就现在的情况，则可以完全跳过去. 因为，由存在与唯一性定理已经知道 $E(t)$ 在整个实数轴 $\mathbf{R}$ 上有定义而且连续. 所以哪怕 $t$ 是一般实数，$E(t)$ 已有定义而且在 $t$ 处连续，根本用不着证明. 何况 $E(t_n)=[E(1)]^{t_n}$ 已经得证. 所以对一般的实数 $t$ 即有

$$\begin{aligned} E(t) &= \lim_{n\to\infty} E(t_n) \qquad \text{(利用 } E(t) \text{ 处处连续)} \\ &= \lim_{n\to\infty}[E(1)]^{t_n} \qquad \text{(上面证明了的)}. \end{aligned}$$

因为 $E(t)$ 在 $\mathbf{R}$ 的一个稠密集 (有理数集) 上与初等方法定义的 $[E(1)]^t$ 相同，因此在整个 $\mathbf{R}$ 上都不妨用 $[E(1)]^t$ 这样

$E(t)$的性质如上面的1，2两点全是利用微分方程方法证明的.

这样一来岂非实数理论等等都是无用而多余的吗？不是.数学中任何一个重要问题总有其关键之处.用不同的解决方法关键点可能不同，但是它们困难的程度都是一样的，甚至这些解决方法本质上都是一样的.上面我们把问题归结为常微分方程解的存在与唯一性定理.而这个定理的证明又少不了证明某个函数序列的一致收敛性.抽象地说来这也是一个完备性问题，与实数理论要解决的问题是一样的，详见第六章度量空间一节.这种情况以后还会出现.

现在余下的问题是，我们很相信 $E(1)=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$，可是怎么证明呢？我们的方法是回到反函数上去，并且用离散的方法处理它.

**2. 增长的离散模型**　现在设自变量 $t$ 不是连续变化的，而是“跳动”着在变，其“步长”是一个有限数 $h>0$，即每一次都是 $t\to t+h$，或写作 $\Delta t=h$，于是 $y(t)$ 也相应有改变量 $\Delta y=y(t+h)-y(t)$.现在设

$$\frac{\Delta y}{\Delta t}=\lambda y,\ y(0)=y_0, \tag{5}$$

问 $y(t)$是什么函数？(5)称为一个差分方程，它是上述增长问题的“离散模型”.人们时常会以为只有连续模型(1)，(2)才是“真实”的，而离散模型则只是近似的.这是不对的.例如 $y(t)$是某人在银行的存款，则后一个条件表明他在期初($t=0$ 时)存入了 $y_0$，利率为 $\lambda$，则 $y(t)$就是他在 $t$ 时的本利和.这个问题本质上就是离散的.由(5)，并以一个存期为单位，即令 $h=1$，有

$$y(1)-y(0)=\lambda y(0),$$

亦即

$$y(1)=(1+\lambda)y(0).$$

所以由 $y(0)$开始会得到

$$y(n)=(1+\lambda)y(n-1)=\cdots=(1+\lambda)^n y(0).$$

如果总的存期仍为1，但把它分成 $n$ 个相等的小间隔，以$\frac{1}{n}$为短的存期.因为存期变短，所以利率也应相应变小，设为$\frac{\lambda}{n}$，仍然以存期的数目的$\frac{1}{n}$为单位，则原来的一个存期现在要算作 $n$ 个小存期.仍用上面的结果可知，在长存期1之末，本利和应该用上式的 $y(n)$来计算而有

$$y(n)=\left(1+\frac{\lambda}{n}\right)^n y(0). \tag{6}$$

不过这时(6)给出的是一个长存期末的本利和，两个长存期末的本利和是 $\left(1+\frac{\lambda}{n}\right)^{2n}y(0)$，…，$t$ 个长存期末的本利和是 $\left(1+\frac{\lambda}{n}\right)^{tn}y(0)$，令 $n\to\infty$，会得到连续的模型.用$\frac{1}{n}$作为时间单位，即有

$$y(t)=\lim_{n\to\infty}\left(1+\frac{\lambda}{n}\right)^{nt}y(0)=e^{\lambda t}y(0). \tag{7}$$

把它和连续模型得到 $E(t)=[E(1)]^t$ 比较，更令人想到 $E(1)$应该就是 $e=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$，但是严格地证明这一点是不容易的.我们不妨来看一看欧拉是怎样来研究这些问题的.首先，我们要给出一个“实际”计算 $E(t)$的方法，为此令 $E(t)=\sum_{n=0}^{\infty}a_n t^n$，则由 $E(0)=1$ 应该有 $a$

由 $E'(t)=E(t)$ 应该有

$$\sum_{n=0}^{\infty} n a_n t^{n-1} = \sum_{n=0}^{\infty} a_n t^n .$$

比较同类项的系数有

$$(n+1)a_{n+1} = a_n，\text{或 } a_{n+1} = \frac{1}{n+1}a_n，$$

所以，$a_{n+1}=\frac{1}{n+1}a_n=\frac{1}{(n+1)n}a_{n-1}=\cdots=\frac{1}{(n+1)!}a_0=\frac{1}{(n+1)!}$，所以

$$E(t) = 1+\frac{t}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^n}{n!}+\cdots. \tag{8}$$

这是一个形式解.当然考虑其收敛性并不困难.所以我们认为已证明了(8)在 $-\infty<t<+\infty$ 上的收敛性以及许可在其上逐项微分、逐项积分.其实对复数 $t$ 也是一样的.详见下一章§4.但是在欧拉的时代，人们对收敛性问题的重要性还很不注意，真正严重地注意到收敛性的数学家应该提到高斯和阿贝尔.即已经是19世纪初年了.欧拉的伟大在于他对数学公式的推演有非凡的才能，对正确的结论有出乎常人的洞察力，虽然自己不甚严格，但得出了许多极为重要的结果.即以 e 的问题为例，他先用二项定理对正整数 $n$ 得出

$$\left(1+\frac{1}{n}\right)^n = 1+\frac{n}{1}\frac{1}{n}+\frac{n(n-1)}{2!}\frac{1}{n^2}+\cdots+\frac{n!}{n!}\frac{1}{n^n},$$

令 $n\to\infty$，欧拉对每一个有限项取极限，则右方的前若干项成为

$$1+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}.$$

再令项数趋于无穷，有

$$\mathrm{e} = 1+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{n!}+\cdots. \tag{9}$$

欧拉的作法从今天看来当然是不能接受的.但是他不但得到了正确的(9)式，而且他用类似的方法处理 $\left(1+\frac{1}{n}\right)^{nt}$，得到了

$$\mathrm{e}^t = 1+\frac{t}{1!}+\frac{t^2}{2!}+\cdots+\frac{t^n}{n!}+\cdots, \tag{10}$$

这与(8)完全一样.所以我们立即有 $E(1)=\mathrm{e}$.可惜，这个证明是不严格的.进一步我们回到 $E(t)$ 的反函数.

**3. 对数函数**　$y=E(t)$ 有反函数吗？上面我们已经证明了 $E(t)$ 是单调函数，它把 $\mathbf{R}$ 单调地(从而是一对一的)、连续可微地映到正实数轴 $\mathbf{R}_+=\{x:x>0\}$ 上.所以，它必有一个反函数，即以 $E(1)$ 为底的对数 $\log_{E(1)} y$.我们把它写成 $t=L(y)$，定义在 $y>0$ 上而且把 $y>0$ 单调地、连续可微地映为 $-\infty<t<+\infty$，而且因为 $t=0$ 时 $y=E(0)=1$，所以 $L(1)=0$.由反函数的导数的定义知道

$$\frac{\mathrm{d}t}{\mathrm{d}y} = \frac{1}{y},\ t\Big|_{y=1} = 0.$$

所以

$$t = L(y) = \int_1^y \frac{\mathrm{d}y}{y}. \tag{11}$$

如果令 $y=1+x$，由上式经变量变换有

$$t = L(1+x) = \int_0^x \frac{\mathrm{d}x}{1+x}. \tag{12}$$

但是，通常的微积分教本告诉我们，(11)恰好就是以 e 为底的对数 $\ln y$，所以

$$L(y) = \ln y. \tag{13}$$

当 $y=E(1)$ 时，由反函数的定义知(11)中的 $t=1$，代入(13)有

$$1 = \ln E(1).$$

从而 $E(1)=\mathrm{e}$. 但是这样做还不能令我们满意，因为我们都熟知

$$\mathrm{e} = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

怎样把 $E(1)$ 与这个极限式联系起来呢？我们已经看到，若用(1)，(2)的离散模型，会见到 $\left(1+\dfrac{1}{n}\right)^n$. 离散模型的“极限”就是连续模型. 这句话一般地说相当有“说服力”，但是总还欠严格. 那么，能否证明 $E(1)$ 直接就是 $\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n$ 呢？这是可以的，我们将在下面讨论.

现在我们要利用积分式(11)直接讨论 $L(y)$ 的性质. 首先是加法定理. 对于 $y_1, y_2\neq 0$，有

$$L(y_1y_2) = \int_1^{y_1y_2}\frac{\mathrm{d}y}{y} = \int_1^{y_1}\frac{\mathrm{d}y}{y} + \int_{y_1}^{y_1y_2}\frac{\mathrm{d}y}{y}.$$

在第二个积分中作变换 $y=y_1\xi$，因为 $y_1\neq 0$，这个变换是合法的，所以有

$$L(y_1y_2) = \int_1^{y_1}\frac{\mathrm{d}y}{y} + \int_1^{y_2}\frac{\mathrm{d}\xi}{\xi} = L(y_1) + L(y_2).$$

这就是对数的加法定理. 本来它由 $L(y)=\log_{E(1)} y$ 以及对数的定义可以直接证明，但是用积分证明似更直接.

其次是两个基本的极限. 在通常的微积分教本中，证明 $\dfrac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^t=\mathrm{e}^t$ 以及 $\dfrac{\mathrm{d}}{\mathrm{d}y}\ln y=\dfrac{1}{y}$ 要用两个基本的极限

$$\lim_{x\to 0}\frac{\mathrm{e}^x-1}{x} = 1,$$

$$\lim_{x\to 0}\frac{\ln(1+x)}{x} = 1.$$

其证明相当不易. 但是用我们现在的方程，这两个不等式都成了几乎不足道的事.

首先我们知道 $E'(t)=E(t)$，而且 $E(0)=1$，所以 $E'(0)=E(0)=1$，由导数的定义立即有

$$E'(0) = \lim_{x\to 0}\frac{E(x)-E(0)}{x} = \lim_{x\to 0}\frac{\mathrm{e}^x-1}{x} = 1,$$

因为我们已经知道了 $E(1)=\mathrm{e}$（尽管还得严格证明），故 $E(x)=[E(1)]^x=\mathrm{e}^x$.

关于第二个极限式，利用 $L(y)=\log_{E(1)} y=\ln y$ 以及(12)式用积分中值定理

$$\ln(1+x) = \int_0^x\frac{\mathrm{d}x}{1+x} = x\Big/(1+\xi),\ 0\leqslant \xi\leqslant x\ .$$

令 $x\to 0$ 即知 $\xi\to 0$，而得

$$\lim_{x\to 0}\frac{\ln(1+x)}{x} = \lim_{\xi\to 0}\frac{1}{1+\xi} = 1.$$

最后是幂级数展开式，对 $E(x)=\mathrm{e}^x$，我们已得到处处收敛的(8)式，对于对数函数，当 $|x|<1$ 时，由于

$$\frac{1}{1+x}=1-x+x^2-x^3+\cdots,$$

逐项积分后，立即有，当 $|x|<1$ 时，

$$\ln(1+x)=\int_0^x\frac{\mathrm{d}x}{1+x}=x-\frac{x^2}{2}+\frac{x^3}{3}+\cdots+(-1)^{n-1}\frac{x^n}{n}+\cdots, \tag{14}$$

这就是尼古拉•麦卡托当年的作法.

**4. 对数的离散处理　$E(1)=\lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n$ 的证明**　从历史上看，对数的出现要早于指数. 上文说的欧拉的贡献都见于他的《无穷小分析引论》一书，出版于1748年，但纳皮尔(John Napier，1550—1617)在1614年就造出了第一本对数表，远早于牛顿的微积分. 另一个瑞士人比尔吉(Jobst Bürgi，1552—1632)也在1620年发表了另一个对数表. 对数的出现确实直接来自航海的需要. 因为航海要作许多乘除法计算，当然人们希望用更简单的加减法来代替它们. 这样就想到了用对数 $y=\log_b t$，它是指数函数 $t=b^y$ 的反函数. 对数的底 $b$，从一开始并不是10，更不是e，而是看如何选取计算最为简单而定. 于是问题就成了给定一定的 $t$ 如何求相应的 $y$，这当然是很困难的. 于是纳皮尔和比尔吉不约而同地想到：把问题反过来做，即给出许许多多的 $y$，再用幂运算来计算 $b^y$，得到许许多多的 $t$. 他们希望这些 $t$ 尽量密集，这样，当你有了一个 $t$ 以后，很有可能这个 $t$ 正是纳皮尔从某一个 $y$ 算出来的，至少很接近纳皮尔算过的值，于是你就得到了相应于这个 $t$ 的 $y$，或者说，就得到了这个 $t$ 的对数，至少是其近似值. 但是用 $b^y$ 算 $t$ 也不简单，除非 $y$ 是整数. 这样，既要 $y$ 是整数，因而 $\Delta y=1$，从而 $y$ 不能太密集，又要求 $t$ 很密集，这个矛盾如何解决呢？办法是取 $b$ 尽可能接近1(但 $b\neq 1$，否则不论 $y$ 是什么值，$t=1^y$ 总是1). 所以比尔吉取 $b=1+10^{-4}$，纳皮尔取 $b=1-10^{-7}$，纳皮尔用小于1的 $b$ 作底是因为航海中需要计算的多是某角的正弦，因而小于1. 如用大于1的 $b$ 作底，对数将是负数，用起来很不方便.

下面只限于讲比尔吉的作法. 我们问当 $y$ 有一个增量 $\Delta y=1$ 时，$\Delta t$ 是多少？$t$ 的相对误差是

$$\frac{\Delta t}{t}=\frac{(1+10^{-4})^{y+1}-(1+10^{-4})^y}{(1+10^{-4})^y}=\frac{1}{10^4}=\frac{\Delta y}{10^4},$$

所以 $\Delta t$ 相当小，而 $t$ 的分布很密. 上式其实就是差分方程

$$\frac{\Delta y}{\Delta t}=\frac{10^4}{t}. \tag{15}$$

如果用 $\dfrac{y}{10^4}=z$ 代替 $y$，则对数的定义式与上述差分方程成为

$$\frac{\Delta z}{\Delta t}=\frac{1}{t},$$

$$t=\left(1+\frac{1}{10^4}\right)^{10^4 z}=\left[\left(1+\frac{1}{n}\right)^n\right]^z, \tag{16}$$

这里 $n=10^4$，如果令 $n\to\infty$，则极限情况是

$$\frac{\mathrm{d}z}{\mathrm{d}t}=\frac{1}{t},$$

$$t = \mathrm{e}^z.$$

这就是 $\mathrm{e} = \lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n$ 的真正的来源.可惜我们不知道欧拉是不是这样想的.但是欧拉对他发现的这个极限似乎情有独钟,所以把自己的姓氏用小写字母写上去了.当然,这不是"证明",只是一种启发,要证明差分方程的解之极限确是微分方程之解,并非易事.证明见后文.

现在我们来求解差分方程,$y = 0$(即 $z = 0$)对应于 $t = 1$,然后用相同的步长 $\Delta y = 1$,即 $\Delta z = \dfrac{1}{10^4}$,从 $z_0 = 0$ 依次得到分点 $z_0 < z_1 < \cdots < z_N$,而相应的 $t$ 的分点是 $t_0 = 1 < t_1 < \cdots < t_N$.$\Delta_i t = t_i - t_{i-1}$ 是不相同的.但这并没有关系,因为由差分方程 $\dfrac{\Delta y}{\Delta t} = \dfrac{10^4}{t}$ 给出 $\dfrac{\Delta z}{\Delta t} = \dfrac{1}{t}$,亦即

$$z_{i+1} - z_i = \frac{1}{t_i}\Delta_i t, \qquad i = 0,1,\cdots,N-1.$$

把它们加起来,并且注意到 $z_0 = 0$,即有差分方程的解

$$z_N = \sum_{i=0}^{N-1}\frac{\Delta_i t}{t_i}. \tag{17}$$

显然它的极限即

$$z = \int_1^t \frac{\mathrm{d}t}{t}.$$

因此,当 $\Delta z \to 0$ 时不但差分方程变成微分方程 $\dfrac{\mathrm{d}L(t)}{\mathrm{d}t} = \dfrac{1}{t}$,而且差分方程的解(17)之极限也正是定义对数函数的积分.

现在我们可以来证明 $E(1) = \lim\limits_{n\to\infty}\left(1+\dfrac{1}{n}\right)^n$ 了.在通常的微积分教本中都要先证明此极限的存在性,并且记之为 e,我们的工作则不是证明 e 的存在(证明见通常的教本),而是承认它的存在,只证明 $\mathrm{e} = E(1)$.

证明的基本思想是从几何着眼.$y = E(t)$ 是微分方程 $\dfrac{\mathrm{d}y}{\mathrm{d}t} = y$ 的解,另外有差分方程(5)

$$\frac{\Delta y}{\Delta t} = y. \tag{18}$$

可是差分是那里算起呢?如图 2 - 1 - 1,在 $y = E(t)$ 上取两点 $P$:$(a, E(a))$,$Q$:$(a+h, E(a+h))$.为了在 $P$、$Q$ 之间考虑逼近 $y = E(t)$,如果要从 $P$ 开始,则用 $F(t)$ 逼近 $E(T)$,这里 $y = F(t)$ 的差分取为:

$$\frac{\Delta y}{\Delta t} = \frac{y - F(a)}{t - a}. \tag{18$_1$}$$

也可以从 $Q$ 开始,用 $G(t)$ 逼近 $E(t)$,而 $y = G(t)$ 的差分取为:

$$\frac{\Delta y}{\Delta t} = \frac{y - G(a+h)}{t - a - h}. \tag{18$_2$}$$

于是在$(18_1)$的情况,差分方程 $\dfrac{\Delta y}{\Delta t} = y$ 成为

$$\frac{\Delta y}{\Delta t}=\frac{y-F(a)}{t-a}=F(a)\text{ 或 }y=F(a)+F(a)(t-a).\tag{$19_1$}$$

在 $(18_2)$ 的情况下则有

$$\frac{\Delta y}{\Delta t}=\frac{y-G(a+h)}{t-a-h}=G(a+h)\text{ 或 }y=G(a+h)+G(a+h)(t-a-h).\tag{$19_2$}$$

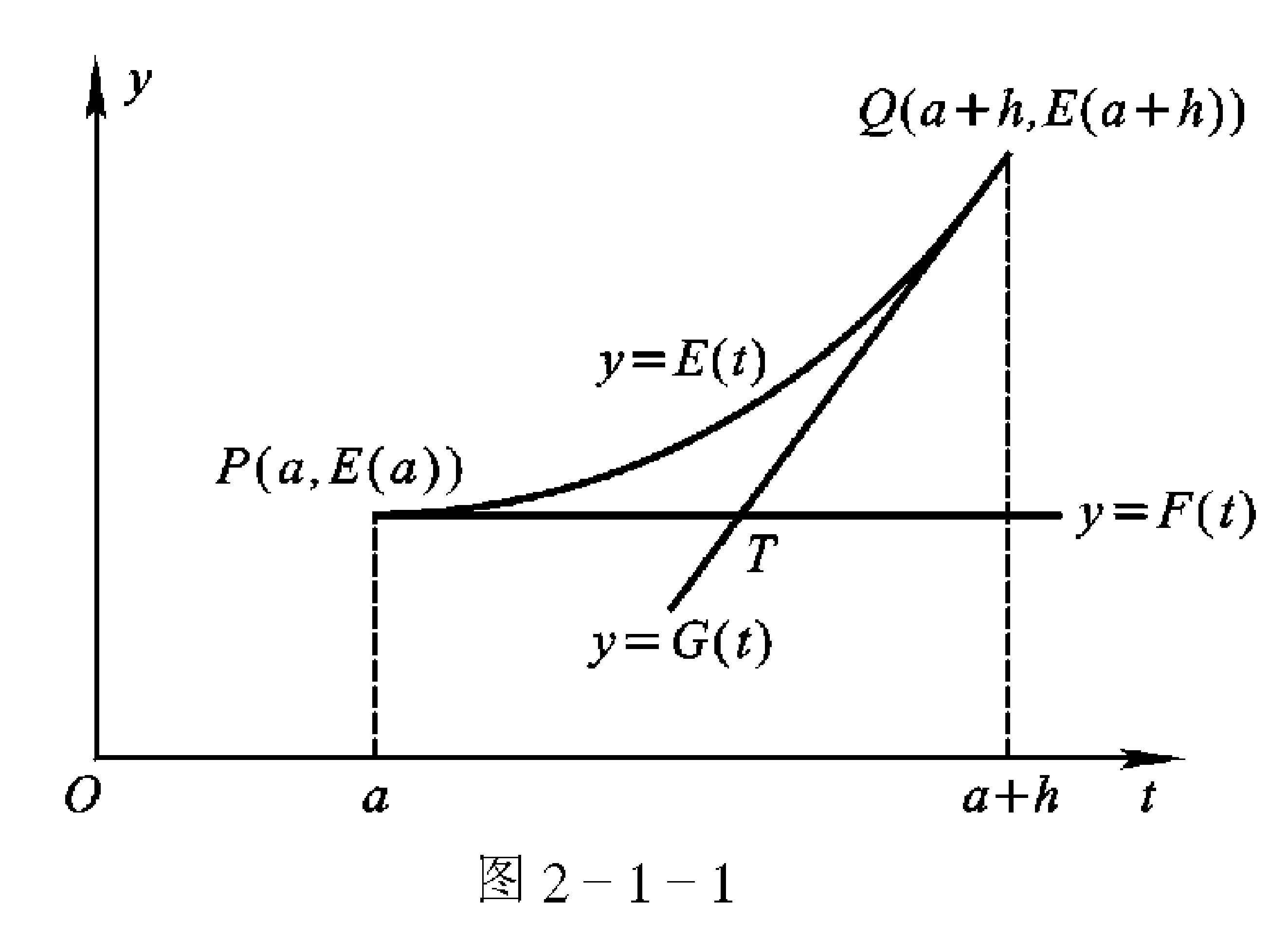



图 2-1-1

注意到 $E'(t)=E(t)$，所以 $E''(t)=E(t)$，但是前面我们已经说过，$E(t)>0$，所以 $y=E(t)$ 是一条向上凹的曲线，因此它必位于 $t=a$ 与 $t=a+h$ 两点之切线 $PT$ 与 $QT$ 之上侧（图 2-1-1），如果我们记 $PT$ 为 $y=F(t)$，$F(t)$ 之表达式即 $(19_1)$，记 $QT$ 为 $y=G(t)$，$G(t)$ 的表达式为 $(19_2)$，所以我们有

$$E(t)\geqslant F(t),\qquad E(t)\geqslant G(t),\tag{20}$$

而且 $E(t)=F(t)$ 只在 $t=a$ 时成立，$E(t)=G(t)$ 只在 $t=a+h$ 时成立.

现在我们把 $t$ 轴上的区间 $[0,1]$ 作 $n$ 等分：$t_0=0<t_1<\cdots<t_n=1$，$t_j=\frac{j}{n}$. 注意到 $E(0)=F(0)=1$，而 $(19_1)$，给出

$$F(a+h)=F(a)+F(a)h=(1+h)F(a).$$

于是令 $a=\frac{i-1}{n}$，$h=\frac{1}{n}$，从 $i=1$ 一直做到 $i=n-1$ 有

$$F(1)=F\left(\frac{n}{n}\right)=\left(1+\frac{1}{n}\right)F\left(\frac{n-1}{n}\right)=\left(1+\frac{1}{n}\right)^2F\left(\frac{n-2}{n}\right)=\cdots$$
$$=\left(1+\frac{1}{n}\right)^nF(0)=\left(1+\frac{1}{n}\right)^nE(0)=\ 1+\ 1$$

而不知道其值究竟是多少.用上式从 $a=\dfrac{n-1}{n}, h=\dfrac{1}{n}$ 开始迭次往前推算,有

$$E(1)=G(1)=\left(1-\frac{1}{n}\right)^{-1}G\left(\frac{n-1}{n}\right)=\cdots=\left(1-\frac{1}{n}\right)^{-n}G(0).$$

$E(t)$ 的图像也在切线 $G(t)$ 之图像上方.现在对 $t=0$ 应用这个事实(上面是在 $t=1$ 处应用 $E(t)>F(t)$ 的),有

$$E(1)=\left(1-\frac{1}{n}\right)^{-n}G(0)<\left(1-\frac{1}{n}\right)^{-n}E(0)=\left(1-\frac{1}{n}\right)^{-n}.$$

综合这两步,得到最后的不等式:

$$\left(1+\frac{1}{n}\right)^{n}<E(1)<\left(1-\frac{1}{n}\right)^{-n}. \tag{21}$$

这就与通常微积分教本中同时考虑两个序列 $\left\{\left(1+\dfrac{1}{n}\right)^{n}\right\}$ 和 $\left\{\left(1-\dfrac{1}{n}\right)^{-n}\right\}$ 完全一致了.不过在通常教材上先证明前一序列上升而有上界,后一序列下降而有下界,因而各有极限,并且证明这两个极限相等,记之为 e.现在我们则是承认这个事实,不过对它作了几何解释,即 $y=E(t)$ 的图像在任一区间 $[a,a+h]$ 上均在其左、右两端的两个切线上方,然后把 $[0,1]$ 分成 $n$ 个等分,并且对每个小区间 $\left[\dfrac{i-1}{n},\dfrac{i}{n}\right]$ 应用上面的事实:或者对左端的 $PT$,我们从左向右,依次令 $i=1,2,\cdots,n$.在左端 $i=1$ 处用 $E(0)=F(0)=1$,而在右端的 $i=n$ 处应用 $E(1)>F(1)$ 得出 $E(1)>\left(1+\dfrac{1}{n}\right)^{n}$.或者对右端的 $QT$ 则由右向左,依次令 $i=n-1,\cdots,0$,并在 $i=n-1$ 处应用 $E(1)=G(1)$,在最左端的 $i=0$ 处应用 $1=E(0)>G(0)$ 而得出 $E(1)<\left(1-\dfrac{1}{n}\right)^{-n}$,总之有(21)式.

注意到

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n}=\mathrm{e},$$

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{-n}=\lim_{n\to\infty}\left(1+\frac{1}{n-1}\right)^{n}=\lim_{n\to\infty}\left(1+\frac{1}{n-1}\right)^{n-1}\lim_{n\to\infty}\left(1+\frac{1}{n-1}\right)=\mathrm{e},$$

立即有

$$E(1)=\mathrm{e}. \tag{22}$$

这就是我们最后的目标.

## §2　周期运动和三角函数

**1. 匀速圆周运动**　周期运动在自然界中几乎是无处不见.天体的运行,不论按日心说或者按地心说,其最简单的"组成成分"都是匀速的圆周运动.正因为日月五星都围绕着一个中心,周而复始,所以研究它们所必需的三角函数,在古希腊数学中占据了很重要的地位.与上一节讨论指数函数与对数函数一样,我们要从微积分学的角度来看待它.但是微积分学既然开始于力学

(其中包括运动学),我们也就从运动学的角度——即只讨论速度与加速度,而不讨论产生加速度的原因——来讨论三角函数.

一切研究工作都应该从最简单的情况开始.牛顿力学始于匀速直线运动,从这里有了惯性系的概念.在讨论周期运动时,我们自然从匀速圆周运动开始.

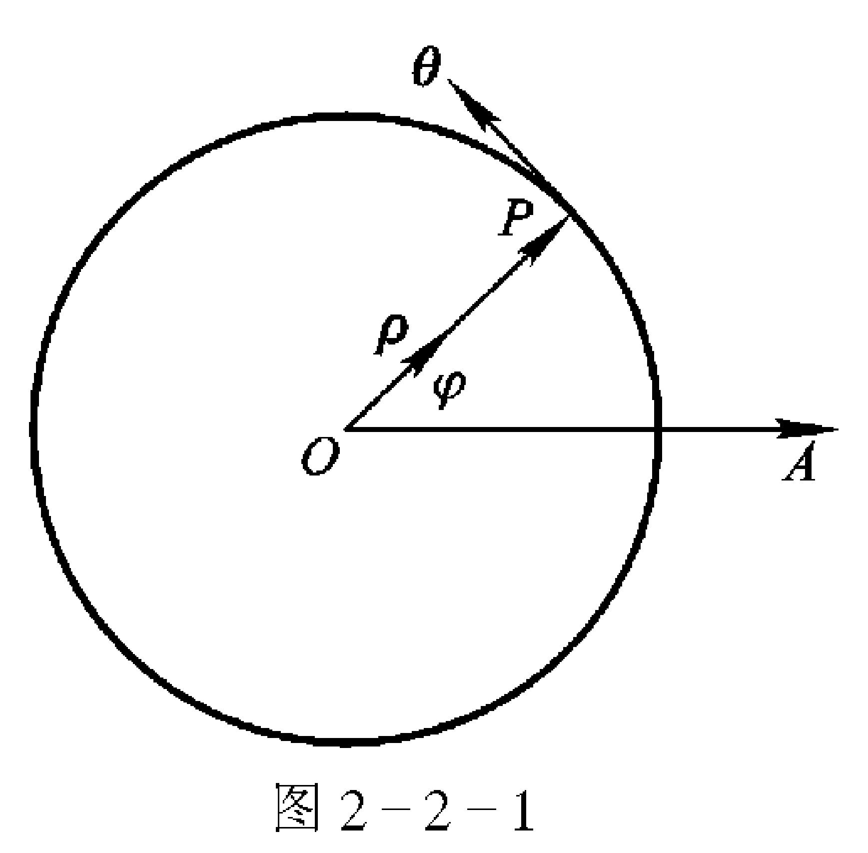


图 2-2-1

表示平面上的运动当然用平面向量为好,圆周运动则用极坐标最方便.以下我们以圆周运动的中心 $O$ 为极坐标原点(如图 2-2-1),以一个指定的轴 $OA$ 为极轴,于是一个点 $P$ 决定了一个向量$\overrightarrow{OP}$.所有的向量均分成两个分量——即轴向的与横向的.在这两个方向上的单位向量记为 $\boldsymbol{\rho}$ 与 $\boldsymbol{\theta}$,一点 $P$ 的辐角 $\varphi$ 不受任何限制:$-\infty<\varphi<+\infty$,所以 $\varphi$ 与 $\varphi+2k\pi$,$k=0,=1,\pm1,\pm2\cdots$ 表示同一个点,这使 $\varphi$ 恰好适用于表示周期现象.而 $\varphi$ 的多值性则时常是产生问题的根源.正如大家都已习惯了的,角度用弧度表示,但是在微积分中弧度有时表示扇形面积占全圆面积的百分比(全圆的面积 $=$ (半径)$^2\times\pi$ 弧度),有时表示角度.但是它总是无量纲量.匀速圆周运动就是角速度 $\alpha=$ 常数的情况.$\alpha$ 的单位是弧度 /秒,但是我们时常以每秒绕圆心的周数 $\omega$ 度量角速度,因此其单位是周 /秒,$\omega=\dfrac{\alpha}{2\pi}$,所以匀速圆周运动的方程是

$$\varphi=\alpha t=2\pi\omega t. \tag{1}$$

后面的表示法是更常用的.在一般地刻画周期运动时,$\alpha$ 称为频率,$\omega$ 称为圆频率.它们的量纲均为$\dfrac{1}{T}$.

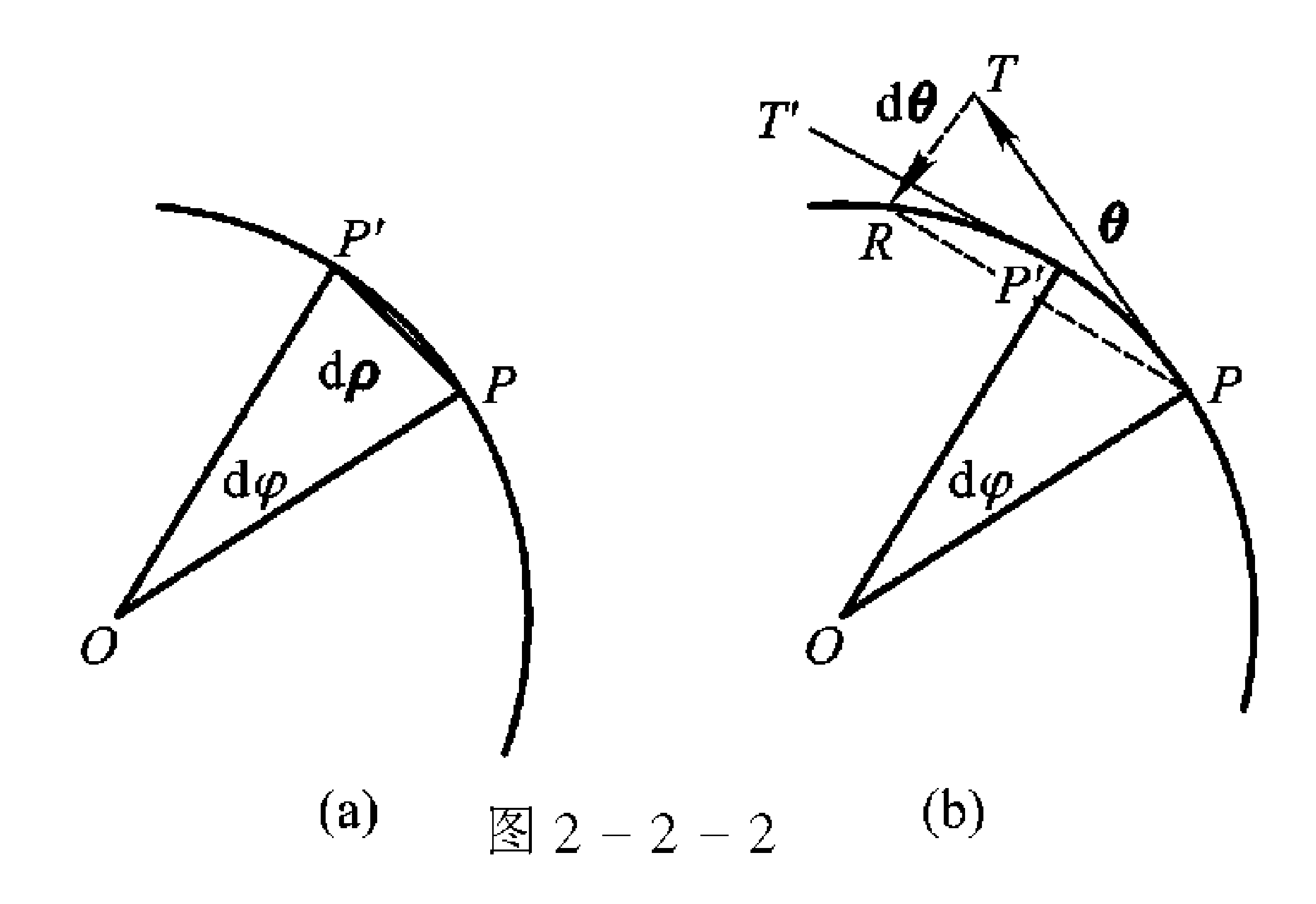


图 2-2-2

下面我们在极坐标中计算运动学的主要变量:速度和加速度,因为我们要在这个基础上建立三角函数理论,所以请读者注意,这里完全不许使用三角函数知识.若 $P$ 点之极坐标为$(r,\varphi)$,注意到随着时间的变化,$\boldsymbol{\rho}$ 与 $\boldsymbol{\theta}$ 大小虽然都不变($|\boldsymbol{\rho}|=|\boldsymbol{\theta}|=1$),但是其方向是改变的,所以,由 $\overrightarrow{OP}=r\boldsymbol{\rho}$,对 $t$ 求导后可以得到

$$\begin{aligned}\boldsymbol{v}&=\frac{\mathrm{d}}{\mathrm{d}t}\overrightarrow{OP}=\frac{\mathrm{d}r}{\mathrm{d}t}\boldsymbol{\rho}+r\frac{\mathrm{d}\boldsymbol{\rho}}{\mathrm{d}t}\\&=\dot{r}\boldsymbol{\rho}+r\dot{\boldsymbol{\rho}},\end{aligned} \tag{2}$$

这里我们按牛顿的作法,用上加一点表示对时间 $t$ 求导.

现在来计算. $\dot{\boldsymbol{\rho}}=\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\rho}$. 如图2－2－2的(a)作单位圆,若 $\boldsymbol{\rho}$ 记为$\overrightarrow{OP}$,经过时间 $\mathrm{d}t$ 后,若 $P$ 沿单位圆变成 $P'$, 则 $\boldsymbol{\rho}$ 有一个增量 $\mathrm{d}\boldsymbol{\rho}$ 即$\overrightarrow{PP'}$, 它对于圆心张一个角 $\mathrm{d}\varphi$, 因此按弧度计, 利用 $|\overrightarrow{OP}|=1$, 即知 $\mathrm{d}\boldsymbol{\rho}$ 之大小即 $\mathrm{d}\varphi$,而其方向则是切线方向$\overline{PP'}$,即 $\boldsymbol{\theta}$ 方向. 总之

$$\dot{\boldsymbol{\rho}}=\frac{\mathrm{d}\boldsymbol{\rho}}{\mathrm{d}t}=\frac{\mathrm{d}\varphi}{\mathrm{d}t}\boldsymbol{\theta}=\dot{\varphi}\,\boldsymbol{\theta}. \tag{3}$$

$\dot{\boldsymbol{\rho}}$ 的方向恰好与 $\boldsymbol{\rho}$ 正交.

再看加速度 $\boldsymbol{a}$. 由(2)

$$\begin{aligned}\boldsymbol{a}&=\dot{\boldsymbol{v}}=\ddot{r}\boldsymbol{\rho}+2\dot{r}\,\dot{\boldsymbol{\rho}}+r\frac{\mathrm{d}\dot{\boldsymbol{\rho}}}{\mathrm{d}t}\\&=\ddot{r}\boldsymbol{\rho}+2\dot{r}\dot{\varphi}\boldsymbol{\theta}+r(\ddot{\varphi}\boldsymbol{\theta}+\dot{\varphi}\dot{\boldsymbol{\theta}}).\end{aligned} \tag{4}$$

所以现在又要计算 $\dot{\boldsymbol{\theta}}$,再看(图2－2－2,b)的单位圆. 若在无穷小时间 $\mathrm{d}t$ 内$\overrightarrow{OP}$ 变成了$\overrightarrow{OP'}$, $\boldsymbol{\theta}$ 由 $P$ 点处切线上的单位向量$\overrightarrow{PT}$ 变成$P'$ 处切线上的单位向量$\overrightarrow{P'T'}$,而 $\mathrm{d}\boldsymbol{\theta}$ 就是向量$\overrightarrow{TT'}$,将 $P'T'$ 平移为 $PR$,则 $|\overrightarrow{TT'}|=|\overrightarrow{PP'}|$. 而由于 $PT\perp OP$, $P'T'\perp OP'$,而且 $|OP|=|PT|=1$,那么, $\triangle POP'\equiv\triangle TPR$. 这样,当 $\mathrm{d}\varphi$ 为无穷小量时, $TT'\perp PT$ 因而指向圆心 $O$ 而且方向相反,这样

$$\dot{\boldsymbol{\theta}}=\frac{\mathrm{d}\boldsymbol{\theta}}{\mathrm{d}t}=-\frac{\mathrm{d}\varphi}{\mathrm{d}t}\boldsymbol{\rho}=-\dot{\varphi}\boldsymbol{\rho}. \tag{5}$$

代入(4)式,即得

$$\begin{aligned}\boldsymbol{a}&=\ddot{r}\boldsymbol{\rho}+(2\dot{r}\,\dot{\varphi}+r\ddot{\varphi})\boldsymbol{\theta}-r\dot{\varphi}^2\boldsymbol{\rho}\\&=(\ddot{r}-r\dot{\varphi}^2)\boldsymbol{\rho}+(2\dot{r}\dot{\varphi}+r\ddot{\varphi})\boldsymbol{\theta}.\end{aligned}$$

在这个计算过程中,我们时常容许一些误差. 但是读者会看到,若以 $\mathrm{d}\varphi$ 表示一阶无穷小量, 们略去的都是关于 $\mathrm{d}\varphi$ 的高阶无穷小量. 在第三章中,我们将指出,略去高阶无穷小量是微分 基本思想,而且全章就是为了说明这个思想怎样由它最初的形态,发展到它的现代的形式, 整个微分学. 所以现在我们就不去讲它了.

(6″)的研究.

**2. 指数函数与三角函数** 平面向量都可以用复数来表示.所以现在我们用 $z=z(t)$ 来表示$\overrightarrow{OP}$,而方程(6″)就成了

$$\ddot{z}(t)+\alpha^2 z(t)=0. \tag{7}$$

如果分开实虚部:$z(t)=x(t)+\sqrt{-1}y(t)$ 就有

$$\ddot{x}(t)+\alpha^2 x(t)=0, \tag{8$_1$}$$

$$\ddot{y}(t)+\alpha^2 y(t)=0. \tag{8$_2$}$$

我们通常用 i 表示 $\sqrt{-1}$,但有的物理学家反对,他们认为 i 通常表示例如电流,所以主张用 j 代替之,但还有物理学家反唇相讥说 j 有时也表示某个物理量,例如电流密度.所以大家都同意,在有可能引起疑义时,就用 $\sqrt{-1}$ 表示虚数单位.不论如何,这样一来,复数进入了.这时我们只要注意一点:把 $\sqrt{-1}$ 当作常数处理,因而可以进入或移出微分与积分号外.这不是一个只图一时方便的约定,而有很深的道理.在讨论线性空间时,我们应该讲明,"系数"在哪一个域(域就是可以进行四则运算的地方)中.是实数域还是复数域,线性空间的基本定理都不会改变(但是欧氏空间相应的对象叫埃尔米特空间).所以,引入复数就意味着现在取复数域作为"基域".基域中的元素就是常数.(7)或者(8)是二阶常微分方程.即令在复数域中其柯西问题之解的存在与唯一性定理仍成立,因为一个复的常微分方程如(7),在分开实虚部后就成为一组两个实的常微分方程如($8_1$),($8_2$).不过这时的初始条件应该是给定初始位置与初速,即给出

$$z(0)=A,\qquad \dot{z}(0)=B. \tag{9}$$

不过 $A$ 与 $B$ 现在都是复数.

为了求解方程(7),我们不妨与§1中求解指数方程$\frac{\mathrm{d}E}{\mathrm{d}t}=\lambda E$ 来比较,在那里我们发现了其解为 $E(t)=C\mathrm{e}^{\lambda t}$,就是说指数函数与其导数为"同类项".既然如此,与其高阶导数自然也是"同类项".所以我们不妨试一下看(7)是否有一解可以写为 $z(t)=C\mathrm{e}^{\lambda t}$?不过 $\lambda$ 与 $C$ 现在都应该是复数.好在上面我们已指出,哪怕复数也可以当成常数处理.故以 $C\mathrm{e}^{\lambda t}$ 代入(7),即得

$$C(\lambda^2+\alpha^2)=0.$$

因此 $C$ 仍为未定常数,而 $\lambda=\pm\sqrt{-1}\alpha$,因此得到(7)的两个解

$$z_1(t)=\mathrm{e}^{\sqrt{-1}\alpha t},\qquad z_2=\mathrm{e}^{-\sqrt{-1}\alpha t}=\bar{z}_1,$$

注意 $\bar{z}_1$ 与 $z_1$ 并不线性相关,即使在复数域中也不是.所以这就给出了(7)的基本解系.而(7)之一切解均可表为

$$z(t)=C_1 z_1(t)+C_2 z_2(t).$$

这里 $C_1$ 与 $C_2$ 是任意复数.只要适当选取 $C_1$,$C_2$ 就能满足任意的初始条件(9).

分开 $z_1(t)$ 之实部虚部并引入新记号,即得

$$z_1(t)=C(t)+\sqrt{-1}S(t).$$

注意到 $z_1(0)=1=1+\sqrt{-1}\cdot 0$,$\dot{z}_1(0)=\sqrt{-1}\alpha=0+\sqrt{-1}\cdot\alpha$,即知

$$C(0)=1,\qquad \dot{C}(0)=0;$$

$$S(0)=0,\qquad \dot{S}(0)=\alpha.$$

因此关于 $C(t)$ 与 $S(t)$ 立即有

$$\begin{cases}\ddot{C}+\alpha^2 C=0,\\ C(0)=1,\dot{C}(0)=0.\end{cases} \tag{$10_1$}$$

$$\begin{cases}\ddot{S}+\alpha^2 S=0,\\ S(0)=0,\dot{S}(0)=\alpha.\end{cases} \tag{$10_2$}$$

我们立刻就会看到

$$C(t)=\cos\alpha t,\qquad S(t)=\sin\alpha t.$$

这是因为：由唯一性定理，自然有

$$e^{\sqrt{-1}\alpha t}=\cos\alpha t+\sqrt{-1}\sin\alpha t. \tag{11}$$

(11) 式是极重要的欧拉公式，而且公认是最漂亮的数学公式之一.我们还想多说一些话.$z(t)=\cos\alpha t+\sqrt{-1}\sin\alpha t$ 为什么成立?我们再细细分析一下.这个式子之成立是我们实际验证的结果，验证中的关键步骤又是证明$\frac{\mathrm{d}\sin\theta}{\mathrm{d}\theta}=\cos\theta$.证明这个求导公式当然不会用到(11)式，因此没有循环论证之嫌.但是它用到了微积分中另一个重要的极限式

$$\lim_{\theta\to 0}\frac{\sin\theta}{\theta}=1. \tag{12}$$

这个极限式与另一个极限式$\lim\limits_{x\to 0}\frac{e^x-1}{x}=1$同时被称为两个最重要的极限式.既然已经有了(11)式，当然就会问，这两个极限式的关系是什么?我们先把话说在前面：(12)式其实与后一极限是同一回事.这一点我们慢慢道来.且先问一下，(12)式在微积分教本中是怎样证的?它的基础是，当 $\theta$ 充分小时(不妨设它为正)

$$\sin\theta<\theta<\tan\theta. \tag{13}$$

这里 $\theta$ 要按弧度计，因为若把弧度按弧长来解释，这个不等式就是

弦长 < 弧长 < 切线长.

但是如果再追究一步，何以知道这个不等式成立就会发现它很不好证.因此在微积分教本中至少在这里是用扇形面积来解释弧度的.把问题归结为面积有很深的含意，我们暂时把它放在一边，而承认(13)式，用 $\sin\theta$(这是正数)通除(13)式，又有

$$1<\frac{\theta}{\sin\theta}<\frac{1}{\cos\theta}.$$

于是只要证最后一项极限($\theta\to 0$ 时)为 1 即可，用倍角公式

$$\begin{aligned}\left|\cos\theta-1\right|&=\left|\cos\left(2\cdot\frac{\theta}{2}\right)-1\right|=\left|\left(1-2\sin^2\frac{\theta}{2}\right)-1\right|\\&=2\sin^2\frac{\theta}{2}<2\cdot\left(\frac{\theta}{2}\right)^2\qquad\left(\text{利用 }\sin\frac{\theta}{2}<\frac{\theta}{2}\right)\end{aligned}$$

故当 $\theta\to 0$ 时，$\cos\theta\to 1$，而基本的极限式(12)得证.

倍角公式的来源是和角公式，例如

$$\cos(\theta_1+\theta_2)=\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2.$$

它的证明当

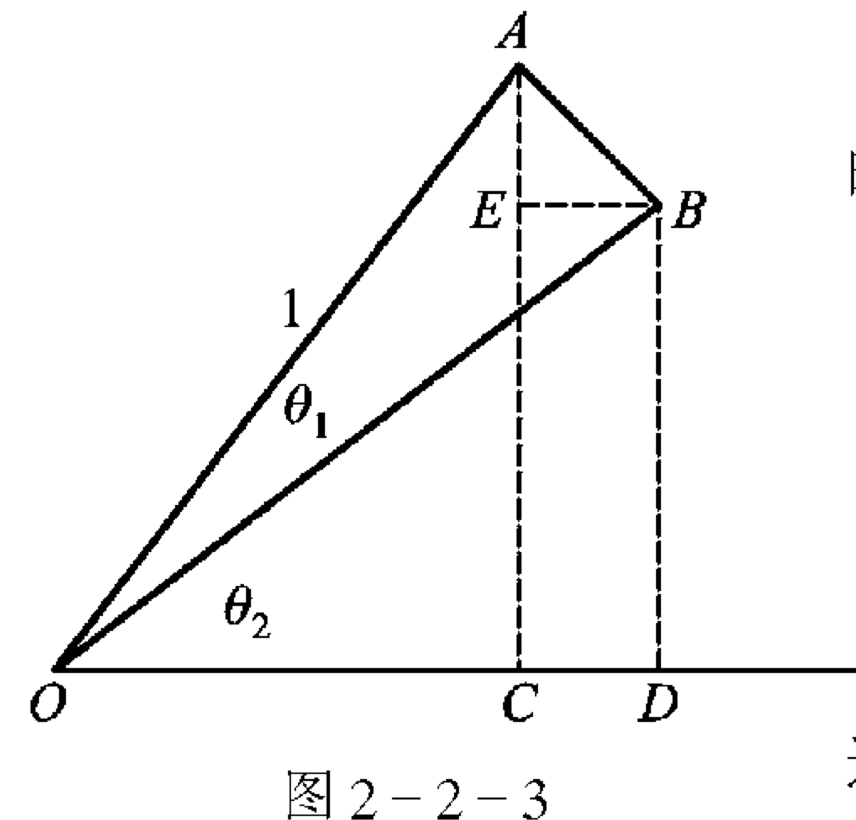


图 2-2-3

$$0 < \theta_1, \theta_2 < \frac{\pi}{2}, \qquad \theta_1 + \theta_2 < \frac{\pi}{2} \tag{14}$$

时确实是十分容易的. 因为如图 2-2-3

$$\begin{aligned}\cos(\theta_1 + \theta_2) &= \frac{OC}{OA} = OC \\ &= OD - CD = OB\cos\theta_2 - AB\sin\angle BAE \\ &= OA\cos\theta_1\cos\theta_2 - OA\sin\theta_2\sin\theta_1 \\ &= \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 \qquad (OA = 1).\end{aligned}$$

这里只要注意到 $\angle BAE = \theta_2$ 即可,但是即使在中学里学三角学时,许多老师以及细心的同学都知道,如果条件(14)不成立,则不但是证明和角公式,甚至画一个适当的图也非易事. 这一切告诉我们,经典的中学教本讲三角函数的办法,不仅谈不上简洁(更不谈优美了),而且相当烦琐,使人难得要领,更不说发展其实质了. 所以上面讲的一切,虽然是对的,我们宁可把它放在一边,进而用其它方法证明 $C(t)$,$S(t)$ 确实是余弦、正弦了. 目前,尽管我们心里明白结果是这样,却不准用这个结果,这样我们需要重新解释

$$e^{\sqrt{-1}t} = C(t) + \sqrt{-1}S(t) \tag{15}$$

(为简单起见,我们取 $\alpha = 1$).

现在我们回到匀速圆周运动,$z(t) = e^{\sqrt{-1}t}$ 就是一个在单位圆周上以匀速 $\alpha$ $(=1)$ 运动的动点,而 $C(t)$,$S(t)$ 正是 $z(t)$(亦即向量 $\overrightarrow{OP}$)在坐标轴上的投影. $\overrightarrow{OP}$ 与 $x$ 轴的交角 $\theta = \alpha t$ $(=t)$ 是 $z$ 的辐角. 这样一来我们就采用了在中学时就应该已经讲过的一般角 $\theta$ $\left(不一定限于 0 < \theta < \frac{\pi}{2}\right)$ 的正弦、余弦定义了:

$$C(t) = \cos t, \qquad S(t) = \sin t.$$

这样做的好处就是我们不必再受(14)式那样的限制. 当然这意味着如和角公式这样简单的东西都得重证. 这样也好,我们可以系统地用(15)式以及 $e^{\sqrt{-1}t}$ 满足方程(7)和初值条件(10)来得出有关正余弦函数的全部性质,以此强调我们是走一条与中学数学不同的路. 我们现在讲法与中学里的讲法最重要的区别何在呢? 把一个向量分解成 $x$ 与 $y$ 轴两个方向的分向量之和,这并没有什么出奇之处. 可是当我们用运动学的思想来看待它时,就发现匀速圆周运动有向心加速度这个简单事实现在成了一个常微分方程(7),而且发现匀速圆周运动可以用 $e^{\sqrt{-1}\alpha t}$ 来表示. 这样向量之分解为分向量就变成了欧拉公式(11). 于是,这个公认的数学中最漂亮的公式从运动学的观点看来,原来简单如此! 因为想法不同,所以还保留记号 $C(t)$ 与 $S(t)$ 一直到最后再改回习惯的 $\cos t$, $\sin t$.

圆具有近乎完美无缺的对称性(直线当然也如此,不过直线可以说是半径为无穷大的圆). 对称性是整个数学的最基本的基础之一. 我们当然现在还不能讲得太玄,但就我们"看得见"的圆的对称性,立即可得出:

$$|z(t)|^2 = C^2(t) + S^2(t) = 1.$$

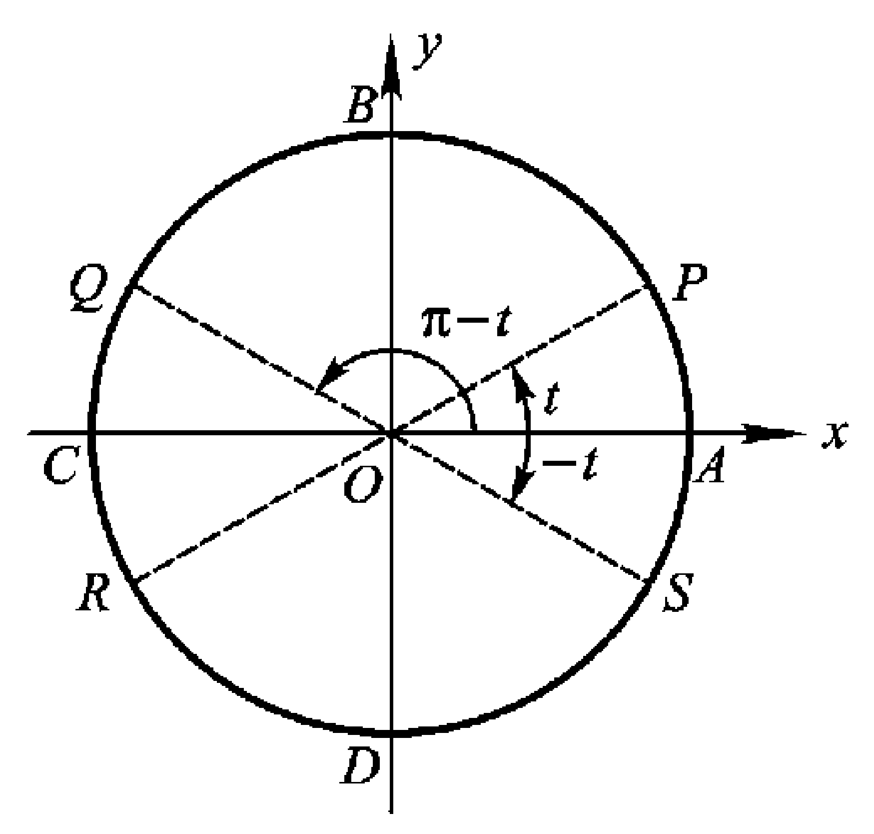


图 2-2-4

当然您可以说这就是勾股定理或者什么别的东西. 要紧的是,不论辐角多少,$z(t)$ 之长总是不变的,这当然是对称性. 其次看"特别角"

的正余弦之值.因为是匀速 $\alpha=1$ 旋转,所以每 $2\pi$ 个单位时间回转一周(所以 $2\pi$ 即周期),又因为旋转是匀速的,所以在$\frac{1}{4}$ 个周期中必转过$\frac{1}{4}$ 个单位圆.这样在 $t=0,\frac{\pi}{2},\pi,\frac{3}{2}\pi,2\pi$ 时 $z(t)$ 位于 $A,B,C,D,A$ 诸点.由对称性和它在 $x$ 或 $y$ 轴上之投影或为 0,或为 1,总之在这些特别角上,依次有

$$e^{\sqrt{-1}\alpha t}=1,\sqrt{-1},-1,-\sqrt{-1},1.$$

再看圆对 $x$ 轴、$y$ 轴(以及原点)之对称性,依次得到 $P$ 之对称点 $Q,R,S$(图 2-2-4),其辐角分别为 $\pi-t,\pi+t,-t$ 由图即见

$$\begin{aligned}&C(\pi-t)=-C(t),\quad C(\pi+t)=-C(t),\quad C(-t)=C(t),\\&S(\pi-t)=S(t),\quad S(\pi+t)=-S(t),\quad S(-t)=-S(t).\end{aligned}\tag{16}$$

(16) 每行中最后的公式即 $C(t)$ 与 $S(t)$ 分别是偶函数与奇函数.最后再加上周期性

$$e^{\sqrt{-1}(t+2k\pi)}=e^{\sqrt{-1}t},\quad C(t+2k\pi)=C(t),\quad S(t+2k\pi)=S(t).\tag{17}$$

以上诸公式我们列为一组.

第二组性质与“相位”(phase)有关.什么叫相位,并没有明确的答案.但是此词是从天文学中借来的.例如月亮的圆缺.阴历初一,天上完全没有月亮,是为朔日,然后有上弦,新月如芽,直到望日(阴历十五)满月,然后是下弦月,终于月亮完全消失,又到了下一个朔望月.这就是月相,是一个周期内月亮处于什么“阶段”——阶段英文也可叫 phase.但是为了讲 $C(t),S(t)$,还是用月亮在 24 小时内的升降——这是由地球自转造成的——来解释更好.如果把图 2-2-4 中的圆看作地球的剖面,我们说 $C(t),S(t)$是在坐标轴上的投影,可是坐标轴的选取有相当大的任意性:$A$ 点的水平轴(指向地平线)是 $y$ 轴,而铅直轴(指向天顶)则是 $x$ 轴,换到 $B$ 点,$B$ 的水平轴恰好是$A$ 的铅直轴,而 $B$ 的铅直轴则是$A$ 的水平轴.因此 $A,B$ 两点的$C(t),S(t)$(各加下标 $A$,$B$ 以示区别)是互换的.如果有人在 $A$ 点,他的时间用 $t$ 表示,于是在 $t=0$(傍晚 6 时)月亮出来了:月出于东山之上,徘徊于斗牛之间,他看见的月亮高度 $S(t)$由 $S(0)=0$ 逐步变成 $S(t)>0$.直到 $t=\frac{\pi}{2}$(半夜 12 时)月正中天,$S\left(\frac{\pi}{2}\right)=1$.如果 $B$ 处还有一个人,他的时间是 $\tau$,则当 $t=\frac{\pi}{2}$ 时,$\tau=0$,于是 $B$ 看见月亮升起到地平线上即$S(0)=0$.$\tau=\frac{\pi}{2}$时,对 $B$ 说来恰好是半夜,而对 $A$ 说来已是$t=\pi$,到了天亮了,月落乌啼霜满天,$S(t)$又变成了 0.总之 $\tau=t-\frac{\pi}{2}$,这时我们说 $t$ 的相位“提前”$\frac{\pi}{2}$,或者由 $t=\tau+\frac{\pi}{2}$,我们也可以说 $\tau$ 有了“滞后”$\frac{\pi}{2}$.由图 2-2-4 看到 $S_A(t)=C_B(t),C_A(t)=-S_B(t)$,所以,以 $A$ 为准(略去下标)

$$S(t)=C\left(t-\frac{\pi}{2}\right),\qquad C(t)=-S\left(t-\frac{\pi}{2}\right).\tag{$18_1$}$$

或者以 $B$ 为准,同样略去下标,又得到

$$S\left(t+\frac{\pi}{2}\right)=C(t),\qquad C\left(t+\frac{\pi}{2}\right)=-S(t).\tag{$18_2$}$$

由此可见,$C(t)$ 和 $S(t)$ 其实是一回事,无非有一个相位差$\frac{\pi}{2}$ 而已.至于是“提前”还是“滞后”,就看对哪一个函数而言.例如对函数 $y=f(t)$,$y=f(t-a)$,$a>0$ 的图像就在其右方相距 $a$

处，所以是 $f(t)$ 的相位"提前"，同样 $y=f(t+a)$，$a>0$，就比 $y=f(t)$ 的相位"滞后". 名词并不重要，重要的是$\frac{\pi}{2}$ 与 $-\frac{\pi}{2}$ 决不可混淆.

把 $(18_1)$ 与 (16) 的最后一个式子 (奇偶性) 合起来，就得到中学生都熟悉的

$$\sin\left(\frac{\pi}{2}-t\right)=\cos t,\qquad \cos\left(\frac{\pi}{2}-t\right)=\sin t.$$

而且在直角三角形中，大家都知道这就是两个锐角互余. 现在我们不限于直角三角形，而且看见一大堆公式，其实都是相位差的问题. 例如

$$\cos(\pi-t)=\cos\left(\frac{\pi}{2}+\frac{\pi}{2}-t\right)=-\sin\left(\frac{\pi}{2}-t\right)=-\cos t,$$

等等，都有多种方法"证明"或"变形". 可是真正重要的是相位问题.

可是相位问题的重要性还不止于此. 例如对于 (3) 式 $\dot{\boldsymbol{\rho}}=\sqrt{-1}\dot{\varphi}\boldsymbol{\theta}$，如果令 $\boldsymbol{\rho}$ 为 $e^{\sqrt{-1}\alpha t}$，其中 $\alpha>0$，$\varphi(t)=\sqrt{-1}\alpha t$，$d\varphi=\sqrt{-1}\alpha dt$，而用复数表示，(3) 就成为

$$\begin{aligned}\frac{d}{dt}e^{\sqrt{-1}\alpha t}&=\sqrt{-1}\alpha e^{\sqrt{-1}\alpha t}=\alpha e^{\sqrt{-1}\frac{\pi}{2}}e^{\sqrt{-1}\alpha t}\\&=\alpha e^{\sqrt{-1}\left(\alpha t+\frac{\pi}{2}\right)}.\end{aligned}\tag{19}$$

所以，$z=e^{\sqrt{-1}\alpha t}$ 求导后，一方面"振幅"放大 $|\alpha|$ 倍，另一方面相位滞后$\frac{\pi}{2}$(如果 $\alpha<0$，就应把它写为 $e^{-\sqrt{-1}|\alpha|t}$，这时仍会得到"振幅"(即模 $|z|$) 放大 $|\alpha|$ 倍，但相位会滞后$\frac{3}{2}\pi$. 与此类似，若对 $z$ 乘以某个复数$\alpha e^{\sqrt{-1}\Psi}$，而 $|\alpha|>0$，则不但有 $z$ 的振幅会放大$\alpha$ 倍，而且也有相位滞后 $\Psi$. 这些现象在物理学中十分重要.

三角函数再一类性质与加法定理有关. 现在我们已经看见了，三角函数通过欧拉公式其实归结为指数函数. 而指数函数最引人注目的性质是加法定理，那么加法定理在三角函数上有什么体现呢? 看 $e^{\sqrt{-1}\varphi_1}$ 与 $e^{\sqrt{-1}\varphi_2}$. 一方面由欧拉公式

$$e^{\sqrt{-1}(\varphi_1+\varphi_2)}=C(\varphi_1+\varphi_2)+\sqrt{-1}S(\varphi_1+\varphi_2).$$

另一方面由指数函数的加法定理有

$$e^{\sqrt{-1}(\varphi_1+\varphi_2)}=e^{\sqrt{-1}\varphi_1}\cdot e^{\sqrt{-1}\varphi_2}=[C(\varphi_1)+\sqrt{-1}S(\varphi_1)][C(\varphi_2)+\sqrt{-1}S(\varphi_2)].$$

所以这两个式子双方必相等. 分开其实、虚部立即得到三角函数的加法定理

$$C(\varphi_1+\varphi_2)=C(\varphi_1)C(\varphi_2)-S(\varphi_1)S(\varphi_2),\tag{$20_1$}$$

$$S(\varphi_1+\varphi_2)=S(\varphi_1)C(\varphi_2)+C(\varphi_1)S(\varphi_2).\tag{$20_2$}$$

有了加法定理自然也就有了倍角公式、半角公式等等.

现在我们看到一个现象: 同一个结果可以用多种不同方法证明. 条条大路通罗马，使人简直不知所从了. 例如，$\frac{d}{dt}C(t)=-S(t)$，$\frac{d}{dt}S(t)=C(t)$，就有多种方法证明: 用指数函数的求导公式加上欧拉公式:

$$\frac{d}{dt}e^{\sqrt{-1}t}=\sqrt{-1}e^{\sqrt{-1}t}=-S(t)+\sqrt{-1}C(t),$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\sqrt{-1}t}=\frac{\mathrm{d}}{\mathrm{d}t}(C(t)+\sqrt{-1}S(t))=\frac{\mathrm{d}}{\mathrm{d}t}C(t)+\sqrt{-1}\,\frac{\mathrm{d}}{\mathrm{d}t}S(t).$$

比较双方实虚部也可以得到同样结果.

我们还可以用相位滞后加上欧拉公式来证.因为上面我们已经得到了指数函数求导后相位滞后$\frac{\pi}{2}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\sqrt{-1}t}=\mathrm{e}^{\sqrt{-1}\left(t+\frac{\pi}{2}\right)}=C\left(t+\frac{\pi}{2}\right)+\sqrt{-1}S\left(t+\frac{\pi}{2}\right).$$

另一方面又有三角函数的相位滞后公式(18):$C\left(t+\frac{\pi}{2}\right)=-S(t)$,$S\left(t+\frac{\pi}{2}\right)=C(t)$,代入上式又一次得到$\frac{\mathrm{d}}{\mathrm{d}t}C(t)=-S(t)$,$\frac{\mathrm{d}}{\mathrm{d}t}S(t)=C(t)$.

同一个结果可以用多种不同方法得到,这当然不是偶然的事.这后面是否还隐藏着更深刻的东西呢?是的,本书第五章全章就是讨论这里的问题.

总结以上的讨论就可以看到,我们的出发点就是指数函数是一个微分方程的解.从运动学的角度来看待这些问题之所以取得成效,也是因为匀速圆周运动的向心加速度公式就是$z(t)$所应满足的微分方程.所以我们是通过$\mathrm{e}^{\sqrt{-1}t}$所满足的微分方程得出了三角函数的性质而完全没有利用直角三角形的边角关系.但是$C(t)$,$S(t)$也直接地是常微分方程柯西问题(10)的解(我们为简单计是限于$\alpha=1$的情况),能否直接由它(仍令$\alpha=1$)得出$C(t)$,$S(t)$的性质呢?这里又遇到了上面说的“条条道路通罗马”的状况:上面讲的许多性质都可以这样得出,但为此我们先把(10)化成一个方程组讨论起来更容易.以($10_2$)为例,令$\dot{S}(t)=\overline{C}(t)$,读者会感到,这不是利用到$\frac{\mathrm{d}S(t)}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\sin t=\cos t=C(t)$吗?何必又设它是$\overline{C}(t)$呢?注意,前面证明$\frac{\mathrm{d}S}{\mathrm{d}t}=C(t)$用的是相位和指数函数的微分方程,我们现在则想用(10),所以暂时还不知道$\dot{S}(t)$是什么,只好写为$\overline{C}(t)$,于是有

$$\dot{\overline{C}}(t)=\ddot{S}(t)=-S(t),$$

$$\ddot{\overline{C}}(t)=-\dot{S}(t)=-\overline{C}(t).$$

利用它们一方面得出$\ddot{\overline{C}}(t)+\overline{C}(t)=0$,另一方面,令$t=0$又得$\overline{C}(0)=1$,$\dot{\overline{C}}(0)=0$,总之$\overline{C}(t)$是以下柯西问题之解:

$$\ddot{\overline{C}}(t)+\overline{C}(t)=0,$$

$$\overline{C}(0)=1,\qquad \dot{\overline{C}}(0)=0.$$

但这就是($10_1$)(令$\alpha=1$),于是由唯一性定理知$\overline{C}(t)=C(t)$这也算是$\frac{\mathrm{d}}{\mathrm{d}t}\sin t=\cos t$的“另证”吧!

总之我们有了一个常微分方程组

$$\begin{aligned}\dot{C}(t)&=-S(t), & \dot{S}(t)&=C(t),\\ C(0)&=1, & S(0)&=0.\end{aligned}\tag{21}$$

有了它，许多性质的证明都十分容易了，例如 $C^2(t)+S^2(t)=1$ 的证明如下：

$$\frac{\mathrm{d}}{\mathrm{d}t}[C^2(t)+S^2(t)]=2[C(t)\dot{C}(t)+S(t)\dot{S}(t)]$$
$$=2[-C(t)S(t)+C(t)S(t)]=0.$$

所以，$C^2(t)+S^2(t)$ 之值不随 $t$ 变化，因此

$$C^2(t)+S^2(t)=C^2(0)+S^2(0)=1.$$

又如加法定理的证明，注意到

$$\frac{\mathrm{d}}{\mathrm{d}t}[S(t)C(a-t)+C(t)S(a-t)]=C(t)C(a-t)+S(t)S(a-t)$$
$$-S(t)S(a-t)-C(t)C(a-t)=0.$$

所以 $S(t)C(a-t)+C(t)S(a-t)=S(0)C(a-0)+C(0)S(a-0)=S(a)$. 令 $a=x+y$，$t=x$，即得

$$S(x+y)=S(x)C(y)+C(x)S(y).$$

这样"证"下去，不免令人感到烦琐而无新意，但是下面我们就立刻会得到引人注目的结果了. 上一节我们给出了指数函数 $\mathrm{e}^t$ 的幂级数展开式

$$\mathrm{e}^t=\sum_{n=0}^{\infty}\frac{t^n}{n!},$$

如果把 $t$ 改成$\sqrt{-1}t$，然后在式右分开实部与虚部，将有

$$\mathrm{e}^{\sqrt{-1}t}=C(t)+\sqrt{-1}S(t)$$
$$=\left(1-\frac{t^2}{2!}+\frac{t^4}{4!}-\cdots\right)+\sqrt{-1}\left(t-\frac{t^2}{3!}+\frac{t^5}{5!}-\cdots\right).$$

所以

$$S(t)=t-\frac{t^3}{3!}+\frac{t^5}{5!}-\cdots=\sum_{n=0}^{\infty}(-1)^n\frac{t^{2n+1}}{(2n+1)!},\tag{$22_1$}$$

$$C(t)=1-\frac{t^2}{2!}+\frac{t^4}{4!}-\cdots=\sum_{n=0}^{\infty}(-1)^n\frac{t^{2n}}{(2n)!}.\tag{$22_2$}$$

这两个幂级数对一切 $t$（不论实数还是复数）都是对的，而且可以逐项微分、逐项积分.

我们也不妨直接把 ($22_1$) 与 ($22_2$) 代入方程组 (21) 及相应的初始条件就知道这两个级数确实适合它们.

现在要回答遗留下来的一个问题：不利用直角三角形边角关系如何证明$\lim\limits_{x\to0}\dfrac{\sin x}{x}=1$?一方面可以说它太容易，不足道了；用级数 ($22_1$) 立即可知，但是这个证明显然"不好"，因为它没有给我们任何新的启发，所以我们来看另一条"道路". 微积分中另一个极重要的极限式$\lim\limits_{x\to0}\dfrac{\mathrm{e}^t-1}{t}=1$ 是怎么证的?当时我们利用 $\mathrm{e}^0=1$，可知

$$\frac{\mathrm{e}^t-1}{t}=\frac{\mathrm{e}^{0+t}-\mathrm{e}^0}{t}=\frac{\Delta\mathrm{e}^z}{\Delta z}\bigg|_{z=0,\Delta z=t}.\tag{23}$$

于是这个极限就是$\dfrac{\mathrm{d}\mathrm{e}^z}{\mathrm{d}z}\bigg|_{z=0}=\mathrm{e}^z|_{z=0}=1$. 注意，我们有意引用记号 $z$，它一般是代表复数的. 那么就要问：如果一个函数 $f(z)$ 自变量是复数 $z$，则

$$f'(z_0)=\lim_{\Delta z\to 0}\frac{f(z_0+\Delta z)-f(z_0)}{\Delta z}$$

对不对?应该说对,只要右方的极限存在即可,但是复的 $\Delta z\to 0$ 比之实的 $\Delta t\to 0$ 内含要丰富得多.如果 $\Delta t$ 是实的,则 $\Delta t\to 0$ 无非是从左侧或右侧趋于 0,而当 $\Delta t$ 是复数时,$\Delta t\to 0$ 的方式就是复杂得很难想象了.因此当 $z$ 为复数时,$\frac{\Delta f}{\Delta z}$ 有极限是一个极强的条件,所以我们就要问,$f(z)=e^z$ 有没有导数呢?复自变量函数的求导是一个很深刻的问题,下一章 §4 专门讨论它,而且恰好$\frac{de^z}{dz}=e^z$ 总是对的.于是我们来看(23)式,并设其中 $t$ 是复的.如果 $t$ 之虚部为 0,则 $t\to 0$ 就意味着实变量 $t\to 0$,这时(23)式就成为

$$\lim_{t\to 0}\frac{e^t-1}{t}=1.$$

如果 $t$ 是纯虚的:$t=\sqrt{-1}x$,而实变量 $x\to 0$,则因

$$\frac{e^{\sqrt{-1}x}-1}{\sqrt{-1}x}=\frac{\cos x-1}{\sqrt{-1}x}+\frac{\sin x}{x},$$

所以(23)式分开实虚部就给出两个极限式

$$\lim_{x\to 0}\frac{\cos x-1}{x}=0,$$

$$\lim_{x\to 0}\frac{\sin x}{x}=1.$$

于是我们看到,微积分中两个基本极限其实是同一个极限的两个特殊情况.

但是这算不得一个证明,只是一条“道路”(approach),要把它变成证明,还得补上复自变量函数求导的许多知识,说它是“道路”,因为它引导我们面对这许多新问题.

**3. 反三角函数** 现在我们已经完成了通过微分方程来定义 $C(t)$ 与 $S(t)$ 并研究它们的性质,完全证实了它们就是中学学过的 $\cos t$、$\sin t$.所以下面我们不再使用 $C(t)$,$S(t)$ 这两个记号了.同样$\sqrt{-1}$ 也采用原来熟悉的 i 来表示.

现在我们要讨论正、余弦函数的反函数问题,我们已知

$$\frac{d\sin\theta}{d\theta}=\cos\theta. \tag{24}$$

而当 $-\frac{\pi}{2}\leqslant\theta\leqslant\frac{\pi}{2}$ 时,由 $\cos\theta\geqslant 0$(这一点由 $\cos\theta$ 是$\overrightarrow{OP}$ 在 $x$ 轴上的投影即可看到)知 $\sin\theta$ 是 $\theta$ 的上升函数.它把$\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$映到$[-1,1]$.因此其反函数 $\theta=\theta(x)$ 存在,而且仍然是一个上升的 $C^1$ 函数(即一阶导数为连续的函数),映$[-1,1]$到$\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$.这样,读者当然会问,$\sin\theta$ 在$\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$以外有没有反函数?当然也可以用类似上面的办法来处理,而例如又可得到$[-1,1]\to\left[\left(k-\frac{1}{2}\right)\pi,\left(k+\frac{1}{2}\right)\pi\right]$,$k=0,\pm 1,\pm 2,\cdots$ 的 $C^1$ 函数,它们都应该是 $x=\sin\theta$ 的反函数 $\theta_k=\tilde{\theta}_k(x)$,也都定义在$[-1,1]$上,但是其值一般不在$\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$中,只有 $k=0$ 的一个例外:

$\tilde{\theta}_0(x) = \theta(x)$. 如果我们说有无穷多个反函数,则称 $k = 0$ 的一个 $\tilde{\theta}_0(x) = \theta(x)$ 为其主值,然后可以讨论这许多反函数之间有什么关系?这个关系式其实很简单,它就是

$$\tilde{\theta}_k(x) = k\pi + (-1)^k\tilde{\theta}_0(x).$$

它相应于正弦函数的一个基本关系式

$$\sin[k\pi + (-1)^k\theta] = \sin\theta.$$

但是我们不来讨论这个问题.我们也可以说 $\sin\theta$ 的反函数是一个"多值函数",至于什么叫"多值函数",下一节再谈.现在我们就只讲主值 $\tilde{\theta}_0(x)$.我们习惯地记它为 $\arcsin x$ 而其它一切 $\tilde{\theta}_k(x)$ 则记为 $\mathrm{Arcsin}\, x$.

由(24)即有

$$\frac{\mathrm{d}\theta}{\mathrm{d}x} = \frac{1}{\cos\theta} = \frac{1}{\sqrt{1-x^2}}. \tag{25}$$

这里根号限取正值,因为对于主值 $\theta = \arcsin x$,$\cos\theta \geqslant 0$ 且当 $x = 0$ 时 $\theta = 0$,所以

$$\theta = \arcsin x = \int_0^x \frac{\mathrm{d}x}{\sqrt{1-x^2}}. \tag{26}$$

按道理说,$x$ 作为 $\sin\theta$ 之值就是$[-1,1]$中的数,而 $\theta$ 作为一个角应该以弧度为单位,弧度可以用弧长来定义,但是更好是用扇形面积来定义.如图 2-2-5. $\theta$ 之弧度值 = 2 扇形 $AOP$ 之面积,换一个记号,我们把(26)中的 $x$ 写成 $y$,$y = \sin\theta$.为什么它称为"正弦"?它就是正对着角 $\theta$ 的半弦 $PR$.但是(26)是否扇形面积?这不要紧,我们来看积分 $\int_0^y \sqrt{1-y^2}\,\mathrm{d}y$,它是曲边梯形 $OAPQ$ 的面积,用分部积分法即有

$$\begin{aligned}\int_0^y \sqrt{1-y^2}\,\mathrm{d}y &= y\sqrt{1-y^2}\Big|_0^y + \int_0^y \frac{y^2\,\mathrm{d}y}{\sqrt{1-y^2}}\\ &= y\sqrt{1-y^2} + \int_0^y \frac{(y^2-1)+1}{\sqrt{1-y^2}}\mathrm{d}y\\ &= y\sqrt{1-y^2} - \int_0^y \sqrt{1-y^2}\,\mathrm{d}y + \int_0^y \frac{\mathrm{d}y}{\sqrt{1-y^2}}.\end{aligned}$$

移项后即有

$$\int_0^y \sqrt{1-y^2}\,\mathrm{d}y = \frac{1}{2}y\sqrt{1-y^2} + \frac{1}{2}\int_0^y \frac{\mathrm{d}y}{\sqrt{1-y^2}}.$$

左边是 $OAPQ$ 面积,右方第一项是$\triangle OPQ$之面积,所以

$$\int_0^y \frac{\mathrm{d}y}{\sqrt{1-y^2}} = 2\text{ 扇形 } AOP \text{ 面积} = \theta \text{ 之弧度值}.$$

这样一来,$y = \sin\theta$ 是由 $\theta$ 之弧度值算出正弦值,而 $\theta = \arcsin y = \int_0^y \frac{\mathrm{d}y}{\sqrt{1-y^2}}$ 是由正弦之长(实际上是半弦)$y$ 算出其所对之角 $\theta$ 的弧度(按扇形面积理解)值.

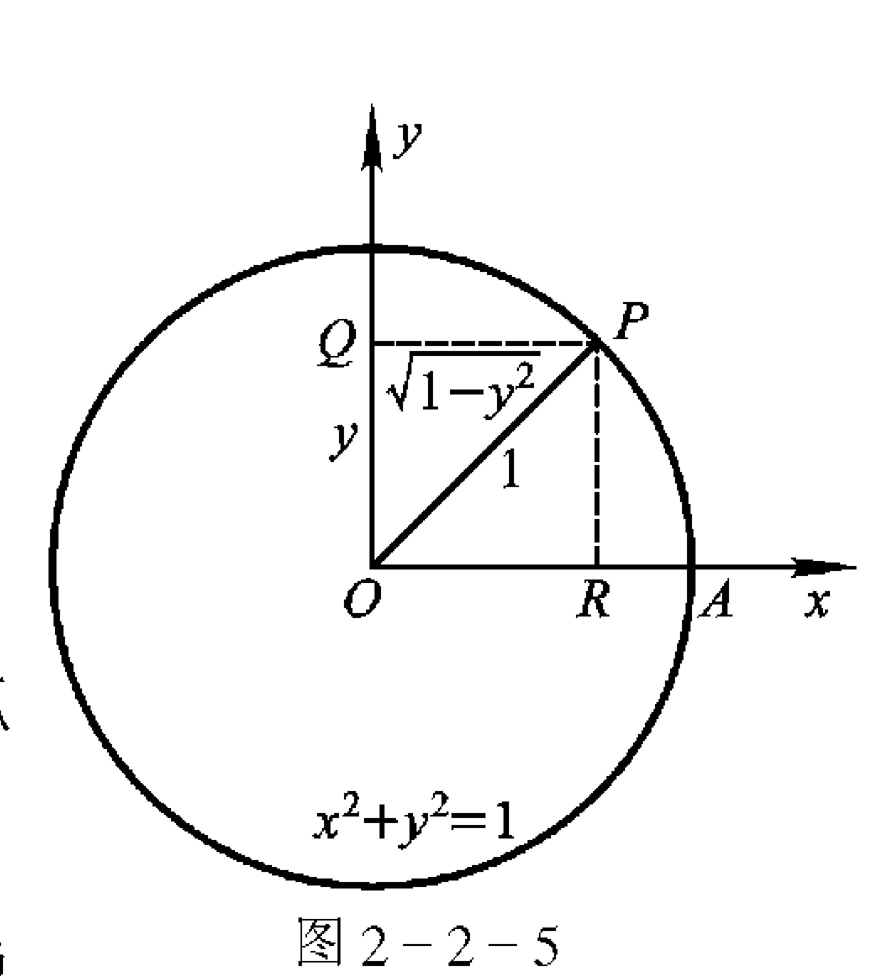


图 2-2-5

用这样的观点重看自然对数 $\eta = \ln\xi$ 更有意义.上节我们说

过，麦卡脱就发现了自然对数 $\eta=\int_1^{\xi}\frac{\mathrm{d}\xi}{\xi}$ 与双曲线的弧下的面积有关. 由图 2－2－6 可见

$$\int_1^{\xi}\frac{\mathrm{d}\xi}{\xi}=\text{曲边梯形 } ABPQ \text{ 之面积}.$$

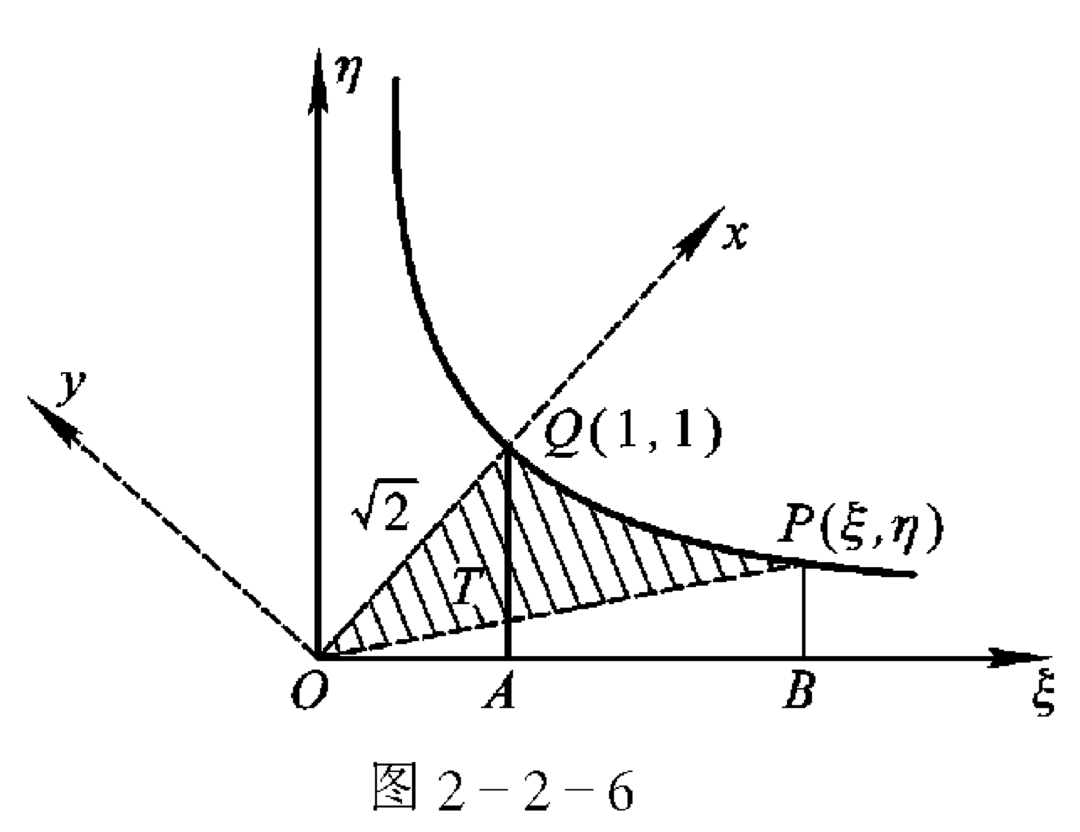


图 2－2－6

但是

$$\triangle OAQ=\frac{1}{2}\cdot 1\cdot 1=\triangle OBP=\frac{1}{2}\xi\cdot\eta=\frac{1}{2},$$

双方同减去 $\triangle OAT$，即有 $\triangle OTQ=$ 梯形 $ABPT$，双方再加上 $TPQ$ 即有

$$\text{扇形 } OPQ \text{ 面积}=\text{曲边梯形 } ABPQ \text{ 之面积}$$

所以，对数函数 $\ln\xi=\int_1^{\xi}\frac{\mathrm{d}\xi}{\xi}$ 作为面积与 $\arcsin x$ 是十分相近的，如果我们作一个旋转

$$\xi=\frac{1}{\sqrt{2}}x+\frac{1}{\sqrt{2}}y,$$

$$\eta=\frac{1}{\sqrt{2}}x-\frac{1}{\sqrt{2}}y,$$

则 $\xi\eta=1$ 变成 $x^2-y^2=2$，则这种相似更为明显. 不过，对于反三角函数

$$\text{扇形 } AOP \text{ 面积}=\frac{1}{2}\int_0^y\frac{\mathrm{d}y}{\sqrt{1-y^2}},$$

式中多了一个因子 $\frac{1}{2}$，是由于图 2－2－5 中的圆是 $x^2+y^2=1$，而图 2－2－6 中的双曲线是 $x^2-y^2=2$. 如果注意到若在圆的方程中把 $y$ 换成 $\mathrm{i}y$，则圆就变成双曲线，所以一旦进入复域，正弦 — 反正弦就会变成指数 — 对数. 欧拉公式也就不足为奇了.

当年麦卡托正是看到了这一点，然后传到牛顿才得到许多初等函数的幂级数展开式. 欧拉在他的《无穷小分析引论》一书中也就把以 e 为底的对数称为“自然对数或双曲对数”.

**4. 振动现象** 以上我们从匀速圆周运动向两个互相垂直的坐标轴投影，找到了 $z(t)$，$C(t)$，以及 $S(t)$ 所适合的常微分方程. 它们的相互关系给出了十分重要的欧拉公式. $z(t)$，$C(t)$，和 $S(t)$ 都是适合同样类型的微分方程

$$\ddot{x}+\alpha^2x=0 \tag{27}$$

(但初始条件不同) 的函数，由于它在物理和工程中的重要性，我们称适合 (27) 的 $x(t)$ 为一个谐振子 (harmonic oscillator). 所以 $z(t)$ 是一个复的谐振子，而 $C(t)$，$S(t)$ 则是实谐振子. 实际用起来，则复谐振子更方便.

一个一般的复谐振子总是 $\mathrm{e}^{\mathrm{i}\alpha t}$ 与 $\overline{\mathrm{e}^{\mathrm{i}\alpha t}}=\mathrm{e}^{-\mathrm{i}\alpha t}$ 的复系数的线性组合. 但是只要弄清楚了其中之一，则另一个自然完全都清楚了. 所以下面我们限于讨论 $C\mathrm{e}^{\mathrm{i}\alpha t}$，这里 $C$ 是一个复数. 但是由欧拉公式我们很容易看到任一复数 $C$ 都可写为 $A\mathrm{e}^{\mathrm{i}\varphi}$，这里 $A\geqslant 0$，所以谐振子一定可以写为

$$A\mathrm{e}^{\mathrm{i}(\alpha t+\varphi)}=A\mathrm{e}^{\mathrm{i}(2\pi\omega t+\varphi)}. \tag{28}$$

这里 $A$ 称为其振幅，$\alpha$ 称为其频率，$\omega$ 称为圆频率，$\alpha$ 之单位就是我们很熟悉的赫兹 (Hertz)，每秒振动一次称为 1 Hz，量纲则是 $\frac{1}{T}$. 所以 $\alpha t$ 或 $\omega t$ 都是无量纲量. $\alpha t+\varphi=2\pi\omega t+\varphi$ 称为相位

(phase)，这样一来，什么叫相位有了一个数学的说法：今后我们时常会遇到形如：$A(x,t)e^{\mathrm{i}\Phi(x,t)}$ 的量，而且 $A(x,t)\geqslant 0$，这时我们总是说 $A(x,t)$ 是振幅，$\Phi(x,t)$ 是相位，$\Phi(x,t)|_{t=0}$ 称为初(始)相(位)．所以对于(28)，$\varphi$ 就是初始相位．

谐振子为什么如此重要?在自然界中我们常见到这样的运动：某一物体(某一质点，某一“电”量，如电位、电容、…)在某一位置附近往复振动．其例子不仅有摆，还有如电子在原子中的振动，所谓脉冲星一阵一阵地例如以X射线形式发出大量的能量，激光在激光器内往复反射……，还有电路中的电流、电压……心电图、血压……这些都是振动．然后这些振动在空间中的传播，就成为波．例如水的振动在水面上传播成为水波，天线中的电磁振动在空间传播为电磁波，电压、电流的振动在导线中传播输送出电讯号．不但自然界如此，社会生活中也有越来越多的现象可以归入其中而用数学方法考察．例如价格、股市、流行病的发生等等都是．尽管研究的对象可以几乎是包罗万象，研究它们的数学方法却是统一而有系统的．常常总是把它们归为一个微分方程(但时常又是一个差分方程)，而(27)可以说是最简单的一例．

(27)是一个理想的情况，称为一维谐振子．说它是一维的，因为刻画这个物理现象只需一个物理量 $x(t)$(时间 $t$ 总是作为参数来看待的)．物理学中许多重要的理论模型都是建筑在它的基础上的．例如考虑单摆(图2－2－7)，就是质量 $m$ 的质点挂在长为 $l$ 的弦上并在重力下振动．要想刻画它只需要知道一个物理量 $x$ 随 $t$ 的变化即可，所以它是一维振动．它是理想的，因为我们假设弦的质量可以忽略不计，而且是刚性的，因此 $l$ 之大小不变．这时质点只受重力 $mg$(向下)作用．重力有两个分力．一是在弦的方向上，因为弦是刚性的，这个分力不起作用．二是在图上的圆弧的切线方向上，大小为 $mg\sin x$ 而且指向平衡位置，所以必与角位移 $x(t)$ 方向相同，但符号相反，称为恢复力．我们考虑单摆的小振动，就是角位移之绝对值 $|x(t)|$ 很小，所以 $\sin x\sim x$，这样恢复力成为 $-mgx$．牛顿的第二定律给出

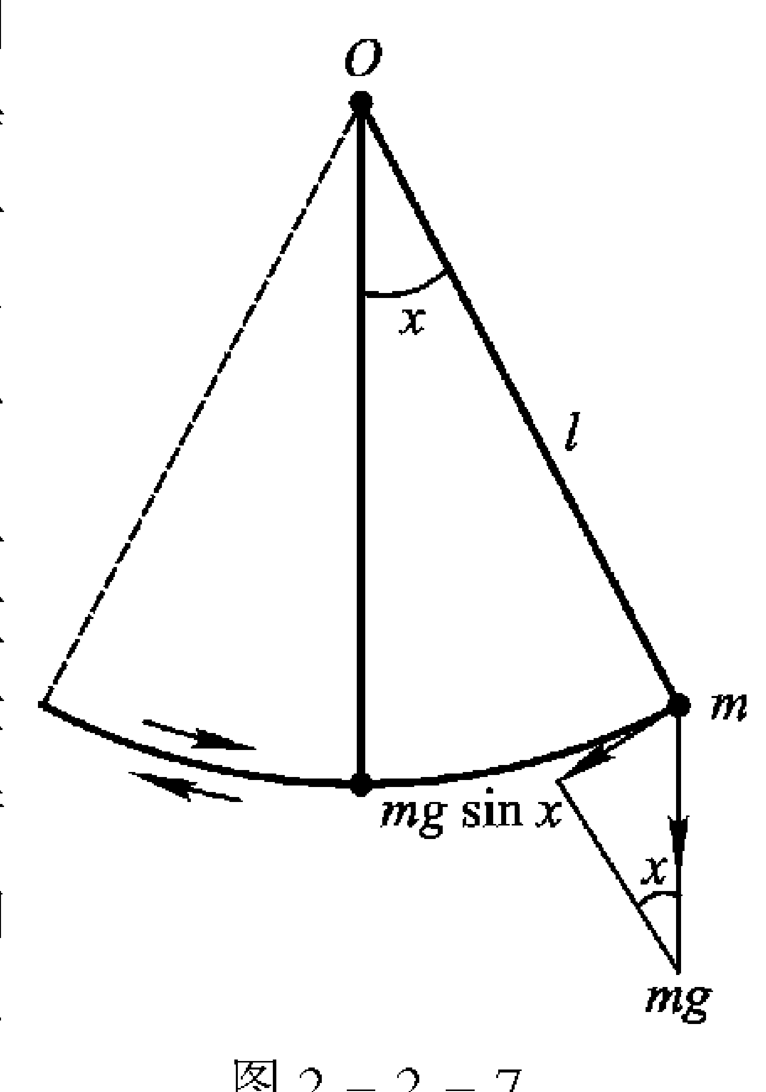


图2－2－7

$$ml\ddot{x}(t)=-mgx(t).$$

左方出现 $l$ 是因为 $lx$ 为线位移，而第二定律的加速度并非角加速度 $\ddot{x}(t)$，而是线加速度 $l\ddot{x}(t)$．此即方程(27)．

上面我们研究三角函数时，是把 $\cos t=C(t)$ 放在 $x$ 轴上，$\sin t=S(t)$ 放在 $y$ 轴上，再考虑一个复平面．$x$ 轴与 $y$ 轴是圆上的点的投影运动的场所，我们不妨称之为物理空间．在单摆的例子中，我们也不妨把角位移 $x(t)$ 放在 $x$ 轴上，这个 $x$ 轴就应称为物理空间(一个恰当的数学名词是构形空间(configuration space)，因为它的一个点 $x$ 就给出了角的大小，因而决定了摆的形状——即构形)，那么是什么相应于 $y$ 轴呢?回到三角函数的情况，$x$ 轴既用来表示 $\cos t$，$y$ 轴则表示 $\sin t=-\dfrac{\mathrm{d}}{\mathrm{d}t}\cos t$．现在在单摆的情况下是否也可以用 $y$ 轴来表示 $\dot{x}(t)$ 即速度呢?(这里相差一个符号，但它显然没有影响．)或者用来表示 $p=m\dot{x}$ 即动量呢?当然都是可以的，真正重要的是要看动量 $p$ 与位移 $x$ 是怎样联系起来的：由方程(27)——其中 $\alpha^2=\dfrac{g}{l}$——双方乘以 $\dot{x}$

再积分，有

$$0=\int_0^t(\dot{x}\ddot{x}+\alpha^2\dot{x}x)\mathrm{d}t=\frac{1}{2}\int_0^t\frac{\mathrm{d}}{\mathrm{d}t}(\dot{x}^2+\alpha^2x^2)\mathrm{d}t$$
$$=\frac{1}{2}(\dot{x}^2(t)+\alpha^2x^2(t))-\frac{1}{2}(\dot{x}^2(0)+\alpha^2x^2(0)).$$

所以 $\dot{x}^2(t)+\alpha^2x^2(t)$ 是守恒的. 但是我们要注意，除了相差一个常数因子 $\left(\text{其中有质量 } m\text{，还有 }\frac{g}{l}\right)$，恢复力是

$$-\frac{\mathrm{d}}{\mathrm{d}x}U(x)=-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\alpha^2x^2}{2}\right),$$

所以 $\frac{\alpha^2x^2}{2}$ 乃是质点的位能，另一方面 $\frac{m\dot{x}^2}{2}$ 是动能，而

$$\frac{p^2}{m^2}+\alpha^2x^2=\dot{x}^2(t)+\alpha^2x^2(t)=\text{常数} \tag{29}$$

就表示能量守恒：

(动能) + (位能) = 常数.

现在我们就作一个平面：$x$ 轴表示位移 $x(t)$，$y$ 轴表示动量 $p(t)=m\dot{x}(t)$，并称此平面为相平面(相空间)(phase plane，phase space) 而与构形空间相区别；$(x,p)$ 就称为此质点的相，因为有此就足以完全决定此质点的运动状况. $x(t)$ 的更高阶的导数不必给了，因为牛顿运动方程就给出了 $\ddot{x}(t)$，而更高阶的导数也可以从运动方程求导数得出. 不但现时该质点的运动状况完全决定了，即其过去未来的运动状况也可以用此时的 $(x,p)$ 作为初始条件再通过方程 (27) 完全决定. 当 $x(t)$，$p(t)$ 已经决定了以后，$(x(t),p(t))$ 就给出了相空间的轨道，称为相轨道，而与构形空间中的轨道 $x=x(t)$ 相区别. 于是对于单摆的小振动，相轨道就是椭圆 (29). 亦即

$$\frac{p^2(t)}{m^2}+\alpha^2x(t)=C.$$

当 $m=\alpha=1$ 时就是圆. 于是，从力学的观点看来，这里的相平面就是前面讲的 $z$ 平面的一般化，而前面讲的

$$\cos^2t+\sin^2t=1,$$

从几何上看无非是勾股定理，而从力学角度来看就是能量守恒定律的特例. 真正值得注意的是，这里得出了椭圆而前面得出了圆，差别仅来自恢复力系数 $\alpha^2$ 之大小. 从拓扑学角度来看，它们并无本质的区别：都是封闭曲线，所以都表示周期运动. 下一章我们要详细讨论开普勒三定律与牛顿万有引力定律的关系，到那时我们会再一次看到行星的运动也包含在这样的框架内.

现在再进一步设质点在阻力亦即摩擦力下运动. 阻力是一个复杂的对象，在最简单的情况下，可以假设阻力为 $-h\dot{x}$，$h>0$ 是一个常数. 这里加了一个负号是因为阻力必与速度方向相反，否则就谈不上阻滞了. 这时由牛顿第二定律将给出

$$\ddot{x}+h\dot{x}+\alpha^2x=0. \tag{30}$$

(我们略去了质量 $m$，而把系数作必要的修正). 我们首先把 (30) 化成一个方程组. 为此令 $x_1=x$，再引入 $x_2=\dot{x}_1$，则 (30) 化成了

$$\dot{x}_1=x_2,$$

$$\dot{x}_2 = -hx_2 - \alpha^2 x_1. \tag{31}$$

$(x_1, x_2)$ 平面现在就是相平面. 所以,研究方程组 (31) 即是在相平面上讨论它,但是为了求解,我们仍用 (30),并且仍设 $x(t) = Ce^{\lambda t}$,代入 (30),即得关于 $\lambda$ 的一个二次方程

$$\lambda^2 + h\lambda + \alpha^2 = 0 \tag{32}$$

称为特征方程,它有两个特征根

$$\lambda_{1,2} = -\frac{h}{2} \pm \frac{1}{2}\sqrt{h^2 - 4\alpha^2}.$$

这时最重要的是 $h^2 - 4\alpha^2$ 的符号. 下面我们只看 $\lambda_1$,因为 $\lambda_2$ 是 $\lambda_1$ 的复共轭,它不会给出新的物理知识.

相应于 $\lambda_1$,将得到解

$$x(t) = Ce^{-\frac{h}{2}t}e^{i\tilde{\alpha}t}. \tag{33}$$

$$\tilde{\alpha}^2 = \alpha^2 - h^2/4.$$

$h = 0$,就是没有阻力的情况,如果 $h$ 逐步增大,把 $C$ 写成一个复数 $C = Ae^{i\varphi_0}$,$A > 0$,有

$$x(t) = Ae^{-\frac{h}{2}t}e^{i(\tilde{\alpha}t + \varphi_0)}.$$

于是我们看到频率由 $\alpha$ 变成 $\tilde{\alpha}$,就是频率变慢. 振幅出现了因子 $e^{-\frac{h}{2}t}$,而当 $t \to \infty$ 时衰减为 0,但是它仍然表现了振动性质. $h = 2\alpha$ 是一个临界值. 这时 $\tilde{\alpha} = 0$. 但是当特征方程出现重根时,解不再写成 (33) 而有变化. 如果阻力再增加,使 $h^2 - 4\alpha^2 > 0$,解 $\lambda_1$ 与 $\lambda_2$ 都成了实数,而解 (33) 现在成了

$$x(t) = Ae^{\lambda_1 t + i\varphi_0},$$

$$\lambda_{1,2} = -\frac{h}{2} \pm \frac{1}{2}\sqrt{h^2 - 4\alpha^2} < 0.$$

这时就不再有振动,而只有指数衰减了.

以上我们是求出了 $x(t)$. 如果转到相空间 $(x_1, x_2)$,尽管 $x_1(t)$,$x_2(t)$ 都不难得出,却不容易找到与椭圆 (29) 相应的联结 $x_1$,$x_2$ 的关系式. 实际上,相轨道不会再是一个封闭曲线,因为不再有周期运动. 但是相轨道的分析引导到数学与物理学中、工程学中十分富有成果的分支.

## §3 进 入 复 域

微积分学在建立了以后,立即就转到复域中了,这个发展是很自然的,又是很重要的. 这样也就形成了数学的一个广阔的领域——复分析. 特别是在 19 世纪,几乎所有重要的数学家都对它有贡献. 许多人都有一个看法,即复分析是当时数学的中心. 其所以如此,一方面来自它对解决实际的物理、工程问题的效用,另一方面则是因为许多在实数域中十分复杂困难的问题在复域中变简单了,其本质更凸现了. 因此复分析就成了一个很完美的框架. 在本书里,我们总是随时准备进入复域,首先讲某一个问题中如何直接应用实域中的微积分理论,一直到遇到了本质上与实域不同之处才止步. 这样为读者将来学习复分析作一个准备.

于是关于函数,读者都知道,狄利克雷关于函数的定义是:设有两个变量 $x$ 与 $y$,$x$ 在某一集

合 $A$ 中变化，而对每个 $x\in A$，均有一个 $y$ 与之相对应，则说 $y$ 是 $x$ 的一个函数(准确些说是 $x$ 的定义于 $A$ 上的函数)，记作 $y=f(x)$，而 $A$ 称为 $f$ 的定义域．这个定义中"一个"两字十分重要，我们下面要仔细讨论．除此以外，有两点值得注意，首先是 $x$ 与 $y$ 也可是复数(但他没有明说)．这样不会造成什么困难，我们只需规定 $x$ 与 $y$ 都是复数即可．然而进一步的研究表明，$x$ 是否为"真正的复数"(即其虚部是否真正为 0)，有极大的关系．以下，凡讲到复变量的函数时，总认为 $x$ 是真正的复数，而不止是作为复数的特例的实数．这时应该更准确地说是复自变量函数．至于 $y$ 是否真正的复数并无关系．即令 $x$ 为实而 $y$ 为复，我们将得到一个取复值的函数．这时不妨分开 $y$ 的实、虚部，并且研究两个取实值的函数．反过来 $x$ 是否真正的复数大有关系，而 $y$ 是否真正的复数则没有关系(只不过在下面讨论的所谓解析函数，当 $x$ 可以取复值时，$y$ 不可能只取实值，除非 $y\equiv$实常数．而在复域中我们又只讨论解析函数)．以后，凡考虑复变量函数时，自变量与函数值总用 $z$ 和 $w$ 记：$w=f(z)$．

另一点值得注意的是，在狄利克雷的定义中 $A$ 是任意集合．但是狄利克雷当时考虑的实际问题中，$A$ 则是一个区间．所以原定义中是说 $x$ 在一区间中变动．就在那时，人们开始认识，有必要考虑任意集合上定义的函数，所以现在的微积分教本中讲函数定义时，定义域总是一般的集合．对复变函数自然也是一样．但是我们要讨论的复变量函数都是解析函数，需要对它进行微分与积分，这时，需要假设 $A$ 为开集，因为要避免在 $A$ 的边界上讨论这些问题．我们常说 $A$ 是一个区域．区域就是连通开集，但有的书上用语不太注意时把开集也就说成了区域．以后在讨论复变量的函数时，我们总规定 $A$ 是区域，整个复平面是一个区域，复平面除去有限多个点也是区域，当然还有更复杂的区域．

有关函数和变量的一些基本概念如极限、连续性、无穷小量以及连续函数的一些基本性质，实的情况与复的情况是一样的．但是因为复数不能比较大小，所以涉及比较大小的概念如单调性、上下确界、最大最小值等等都没有意义．但是 $|f(z)|$ 却是 $z=x+\mathrm{i}y$ 的实值函数，也就是 $x,y$ 的实值函数，所以若 $f(z)$ 在有界闭区域 $G$ 上连续．则 $|f(z)|$ 必可在 $G$ 上达到最大最小值，可以取中间值等等，但是对 $f(z)$ 就谈不上这些问题．在涉及微分积分时，如前面所指出的，只需注意复常数可以如实常数一样看待，可从微分与积分号下提出来或放进去，均与实的情况一样．当然，中值定理因涉及比较大小就不可能成立了．这些问题在下面两章讨论复变量函数的微分与积分时都会详细讨论，这里就一句带过．

我们最要注意的是"多值函数"的问题．按函数的定义，所谓 $y=f(x)$ 是 $x$ 在定义域 $A$ 上的函数，即指对于一个 $x\in A$，必有"一个" $y$ 与之相应．上文我们特别指出"一个"两字的重要性，所以按函数的定义，函数就是单值函数，无所谓多值函数．

这样的限制是有道理的．如果我们允许对同一个 $x$ 有多个 $y$ 与之相应，则下面的"东西"算不算一个多值函数？

$$y=\begin{cases}\sin x\\ \tan x\\ x^3+\mathrm{e}^x\\ \cdots\end{cases}\tag{1}$$

这样把一些毫不相干的东西硬拉在一起并称之为"多值函数"的作法，至少是没有益处的．

但是在数学中给出一个定义就是划了一条界线．在界线内的对象是需要研究的，而在界线外

的如上面的例子就可能是需要排除的.因此界线的划法就很有讲究:应该要能促进数学研究的进展.现在我们要问,在给出了函数的定义如上以后,我们在数学的研究与应用中常用的多值函数例如 $w=\sqrt[n]{z}$ 应如何对待?

我们不要把函数按定义必须是单值函数这一点看死了.在现代数学中有所谓集值映射(set-valued map)的概念,它在例如经济数学中就很有用.它的定义是,对每个 $x\in A$,必有某个"大集合"即一个空间 $Y$ 的子集 $y$(注意现在是 $y\subset Y$,而不是 $y\in Y$)与之对应,$y$ 中可能有许多元素如 $y=\{y_1,y_2,\cdots\}$(甚至不是可数多个),$y_j\in Y$ 而不是 $y_j\subset Y$.这个概念的来源之一就是如同 $w=\sqrt[n]{z}$ 这样的"多值函数".集值映射这个概念常见于"非线性分析"这个很新很活跃的分支,我们不能涉及.只说一下,读完本书对进入这个领域大有好处.

我们现在讨论 $w=\sqrt[n]{z}$ 却是从一个很古老的观点出发的:把它看成 $w=z^n$ 的反函数.这里的记号要解释一下,$w=\sqrt[n]{z}$ 是 $z=w^n$ 的反函数,为什么又说 $w=\sqrt[n]{z}$ 又是 $w=z^n$ 的反函数呢?要注意重要的是"函数关系":"$n$ 次根"函数关系与"$n$ 次幂"函数关系互为反函数.如果某一个函数关系记为 $f$,而且已知有反函数存在,则反函数关系,亦即逆映射记为 $f^{-1}$.于是 $f\cdot f^{-1}=id$,$f^{-1}\cdot f=id$. $id$ 是恒等映射.不过这两个 $id$ 并不相同,因为其定义域不同.前一个 $id$ 定义在 $f$ 的值域上,后一个定义在 $f$ 的定义域上.总之我们现在讨论 $w=\sqrt[n]{z}$.若固定 $z$ 值,它是 $n$ 次代数方程

$$w^n - z = 0 \tag{2}$$

的根.这是一个很古老的问题,当年高斯第一次证明了 $n$ 次代数方程 $\sum_{k=0}^{n} a_{n-k}w^k=0\ (a_0\neq 0)$ 必有 $n$ 个复根存在,这是数学史上第一次有了一个"纯粹的存在定理",不过我们现在要讨论的是更一般的方程

$$a_0(z)w^n + a_1(z)w^{n-1} + \cdots + a_n(z) = 0, \tag{3}$$

而系数 $a_i(z)$ 都是 $z$ 的多项式.从这种方程解出的 $w=w(z)$ 称为代数函数.它们大多是多值的,这个问题在19世纪中后期是数学中的重大问题,黎曼正是由研究它们才得到十分深刻的黎曼曲面的思想.所谓黎曼曲面 $S$ 是这样一种"曲面",使得上述方程的"多值"的解成为 $S$ 上的单值函数.由于(2)是(3)的特例,我们现在就以(2)为特例介绍这个极为深刻的思想.

对于固定的 $z$,方程(2)有 $n$ 个根

$$w^{(k)} = \mathrm{e}^{\mathrm{i}2k\pi/n}, \qquad k = 0,1,\cdots,n-1, \tag{4}$$

如果让 $k$ 取其他整数值如 $k=n,n+1,\cdots$ 或 $k=-1,-2,\cdots$ 所得的 $w^{(k)}$ 仍然在(4)式所给的 $n$ 个 $w^{(k)}$ 之中,(4)中的 $n$ 个 $w^{(k)}$ 称为1的 $n$ 个 $n$ 次单位根,它们位于一个正 $n$ 边形的 $n$ 个顶点上(如图2-3-1),其中一个是1(即 $w^{(0)}$).

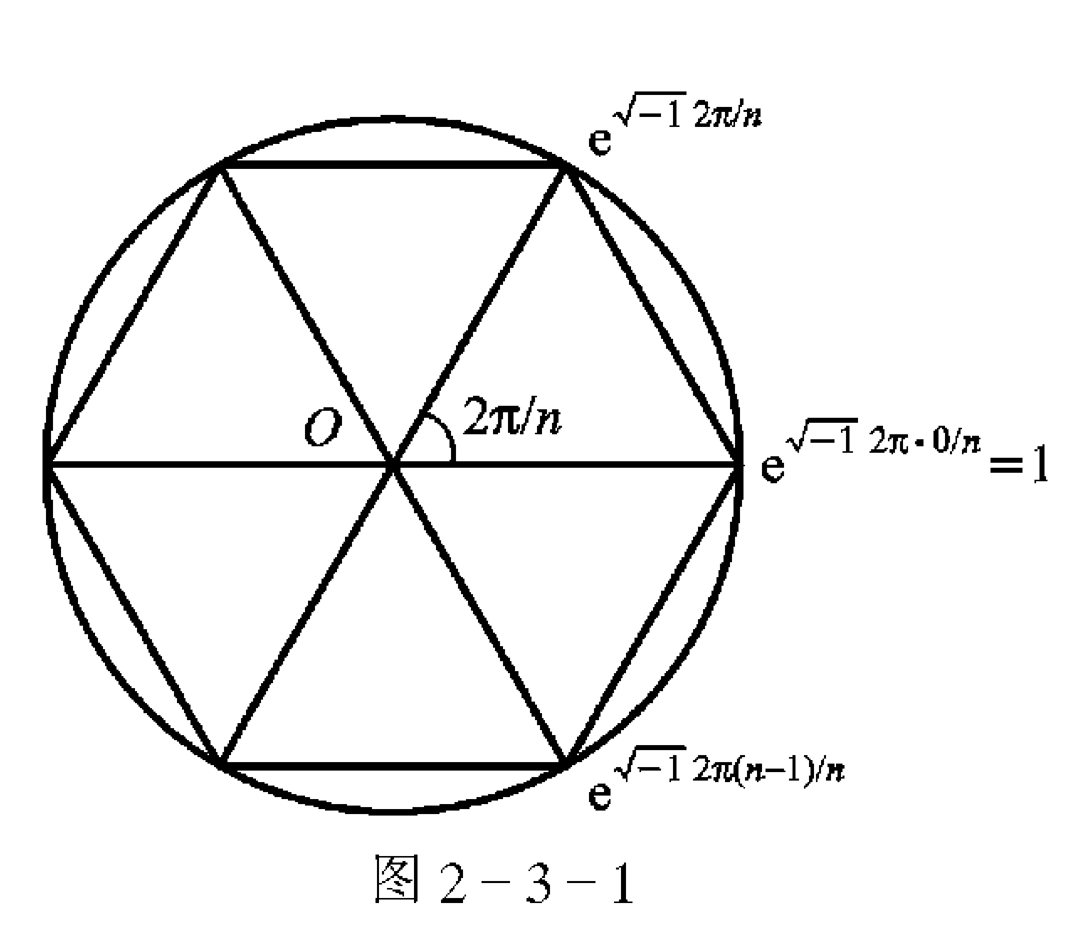


图2-3-1

我们可以再叙述一次上面的结果.由欧拉公式,任何复数 $z$ 均可写为 $z=r\mathrm{e}^{\mathrm{i}\varphi}$,这里 $r\geqslant 0$,$\varphi$ 称为辐角,它是多值的,因为若 $\varphi$ 是一个辐角,则 $\varphi+2k\pi$,$k=0,\pm1,\pm2,\cdots$ 都是辐角.但 $r=0$ 时辐角不确定.容易证明,(2)的 $n$ 个根即

$$w^{(k)} = z^{\frac{1}{n}} = r^{\frac{1}{n}}\mathrm{e}^{\mathrm{i}(\varphi+2k\pi)/n}, \quad k = 0,1,\cdots,n-1. \tag{5}$$

$r^{\frac{1}{n}}$，也就是正实数 $r$ 的 $n$ 次算术根. 如果 $z$ 是一正实数，则可取辐角 $\varphi=0$，这时在上述 $n$ 个根中有一个是正实数(相应于 $k=0$)，这个根也就称为 $z^{\frac{1}{n}}$ 的算术根. 而对于一般的复数，算术根的概念是没有意义的. 有的书上规定只有算术根才可使用 $\sqrt[n]{z}$ 的记号，一般情况下则使用 $z^{\frac{1}{n}}$. 我们却不如此拘谨，主要是要知道它是"多值函数"，而且这种多值性的来源就在于复数辐角的多值性. 为了从这许多值中分出特别的一个便于使用，我们时常指定辐角 $\varphi$ 的取值范围是 $0\leqslant\varphi<2\pi$，并称这个值为"主值". 但是这又是一个不必过分计较的名词，有时也说 $-\pi\leqslant\varphi<\pi$ 是主值、$-\pi<\varphi\leqslant\pi$ 是主值，$-\pi\leqslant\varphi\leqslant\pi$ 是主值……总之，目的在于分离出一个特定的值来.

图 2-3-2

乍看一下，$w^n=z$ 有 $n$ 个反函数 $w^{(k)}$，但仔细看一下又不能泰然处之，因为若令 $z_0=r_0\mathrm{e}^{\mathrm{i}\varphi_0}$，$0\leqslant\varphi_0<2\pi$，并取 $w^{(0)}(z_0)=r_0^{\frac{1}{n}}\mathrm{e}^{\mathrm{i}\varphi_0/n}$ 为 $w^{(0)}(z)$ 之"初始值"($w^{(0)}(z)$ 不妨称为 $w=z^{\frac{1}{n}}$ 的主值)，现在来看 $z$ 变化时 $w$ 之值如何变化. 这时 $z$ 变动的道路(path)之构造将起关键作用. 例如取图 2-3-2 上的不绕过 $O$ 的道路 $L_1$. 如果 $z$ 从 $z_0$ 开始依逆时针方向绕 $L_1$ 一圈回到 $z_0$，则 $z$ 变回原来的 $r_0\mathrm{e}^{\mathrm{i}\varphi_0}$，而 $w$ 也由 $w^{(0)}(z_0)$ 变回原来的 $w^{(0)}(z_0)$. 但是 $z$ 从 $z_0$ 开始沿绕过 $O$ 的道路 $L_2$ 依逆时针方向又回到 $z_0$，后一个 $z_0$ 将是 $r_0\mathrm{e}^{\mathrm{i}(\varphi_0+2\pi)}$，而 $w$ 将由 $w^{(0)}(z_0)$ 变成 $r_0^{\frac{1}{n}}\mathrm{e}^{\mathrm{i}(\varphi_0+2\pi)/n}$，即变为 $w^{(1)}(z_0)$. 若 $z$ 再沿 $L_2$ 绕 $O$ 一圈再次回到 $z_0$，现在的 $z_0$ 将是 $r_0\mathrm{e}^{\mathrm{i}(\varphi_0+2\cdot 2\pi)}$，而 $w$ 由 $w^{(1)}(z_0)$ 变成 $w^{(2)}(z_0)=r_0^{\frac{1}{n}}\mathrm{e}^{\mathrm{i}(\varphi_0+2\cdot 2\pi)/n}$. 所以这 $n$ 个 $w=z^{\frac{1}{n}}$ 将依次互相转变. 所以把 $w=z^{\frac{1}{n}}$ 看成一个单值的反函数更为妥当. 但是现在给定一个 $z$，可以求出 $n$ 个 $w$ 值如(5)，这个矛盾如何解决？办法就是把原来的 $z_0$ 与绕 $L_2$ 转了一圈使 $\varphi_0$ 增加了 $2\pi$ 的 $z_0=r_0\mathrm{e}^{\mathrm{i}(\varphi_0+2\pi)}$ 区别开来. 这里就表现了黎曼的天才：设想有 $n$ 个 $z$ 平面，如同 $n$ 张纸一样，而且把它们钉在一起，可是不是像书那样沿书脊钉，而且在 $z=0$ 处钉住，然后例如把它们都沿正实轴剪开，再把一张纸的沿正实轴的下沿(辐角接近 $2\pi$ 处)与另一张纸的上沿(辐角接近 0 处)粘在一起，像一个旋转楼梯. 当我们沿楼梯旋转一周环绕 $O$ 点就自动上了一层楼. 初始的 $z_0$ 在一楼，下一个 $z_0$ 在二楼……这样就把两个 $z_0$ 区别开来了. 这样一层又一层地转，$w^{(0)}$ 就自动地变成 $w^{(1)}$……再变成 $w^{(n)}$ 等等. 可是到了 $n$ 楼(即 $k=n-1$ 处)，如果再转，$w^{(n-1)}$ 就应变成 $r_0^{\frac{1}{n}}\mathrm{e}^{\mathrm{i}(\varphi_0+2(n-1)\pi+2\pi)/n}=r_0^{\frac{1}{n}}\mathrm{e}^{\mathrm{i}(\varphi_0+2n\pi)/n}=r_0^{\frac{1}{n}}\mathrm{e}^{\mathrm{i}\varphi_0/n}$，所以 $z_0$ 就应该回到一楼的起点(图 2-3-3). 可是 $z_0$ 怎么能由 $n$ 楼神不知鬼不觉地再回起点呢？$z_0$ 的路径应该与其它楼面都相交，可是交点处的 $z_0$ 对应的 $w$ 值又怎么计算呢？可是黎曼认为不必管，就让 $z$ 像崂山道士那样钻天入地了无痕迹算了，黎曼把这样造出来的东西 $S_n$ 看成一个曲面，后世就称为黎曼曲面. 于是所谓多值函数 $w=z^{1/n}$ 就成了黎曼曲面 $S_n$ 上的单值函数. 自此以后，研究多值函数就变成了研究如何构造相应的黎曼曲面.

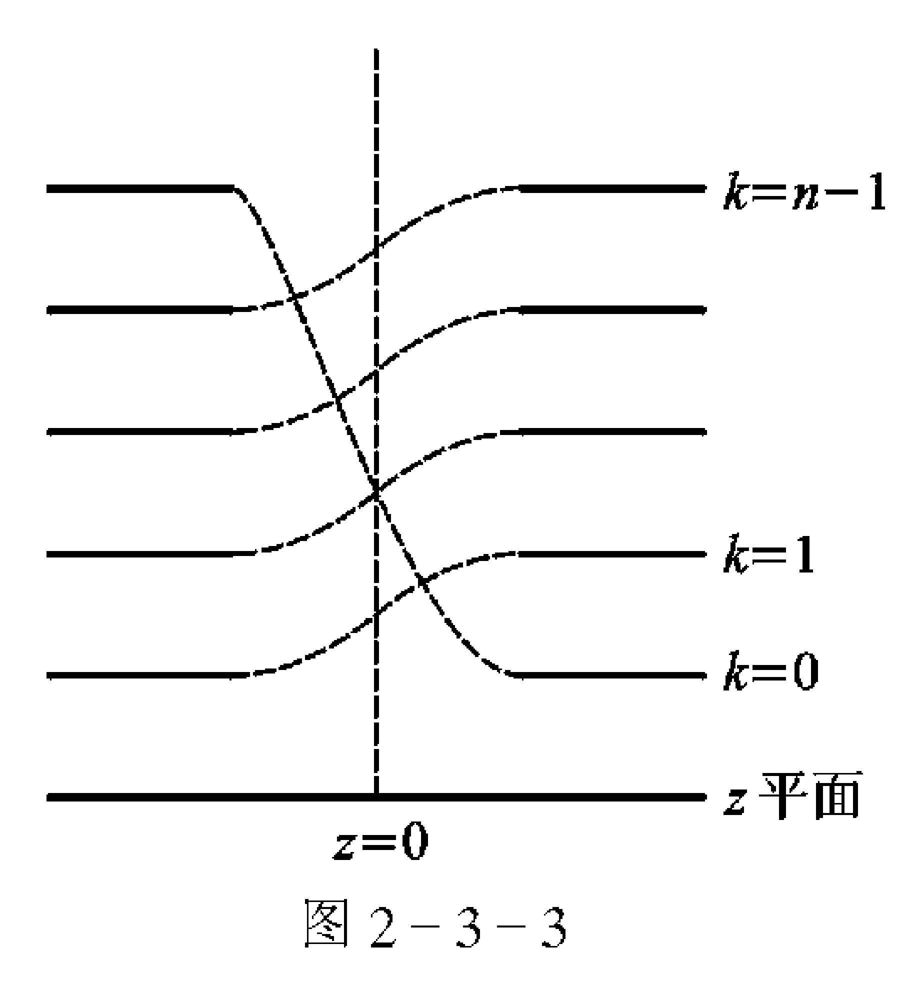


图 2-3-3

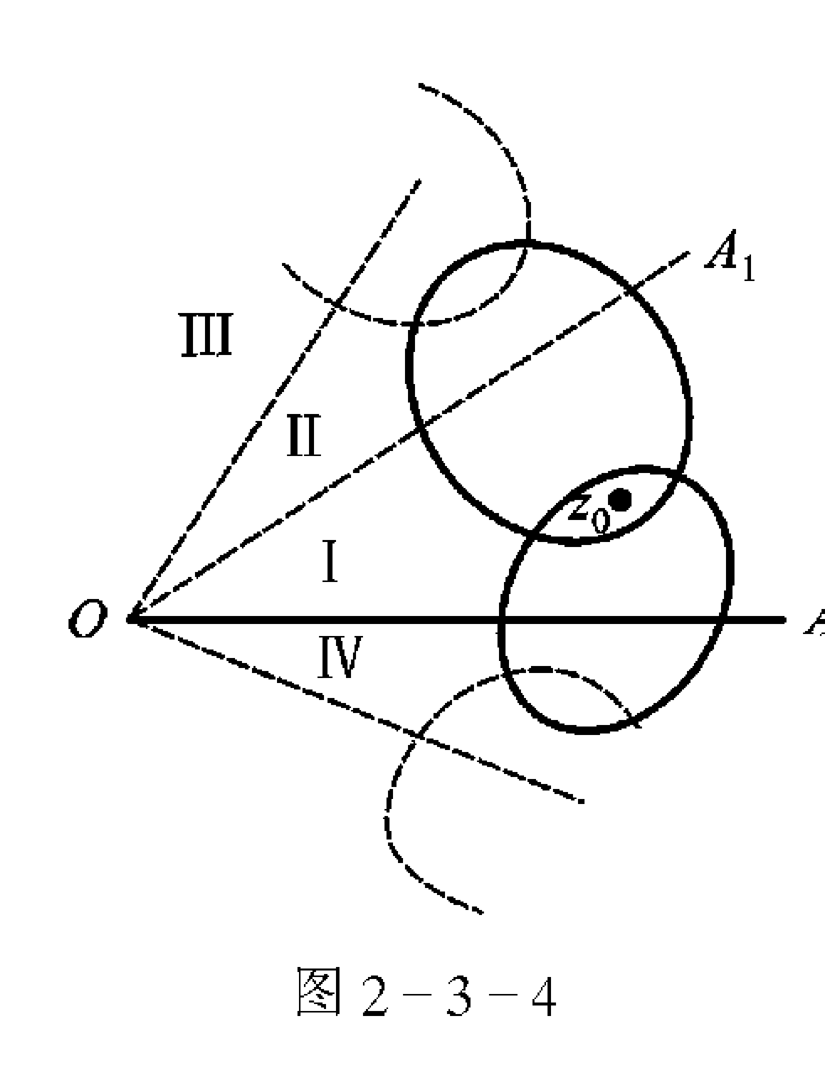



图 2-3-4

当然，读者会感到困惑，$S_n$ 究竟是怎样构造的呢？我们不妨这样来作.先找一个小纸片，在上面标记一点 $z_0$，然后例如沿着 $L_2$ 与 $z_0$ 相近处，再找一个小纸片与第一片粘起来，粘了一整圈以后，当 $L_2$ 回到 $z_0$ 时，注意在这时的小纸片上虽然仍然标记上 $z_0$，可是不要与原来最初的小纸片粘在一起，而要用新的小纸片粘下去，这样到了第 $n$ 圈时，已经用过了不少小纸片.后来的小纸片与第二片、第三片大相径庭，所以可以“置之不理”，而硬是把它与最初的一片粘在一起.这有点像运动会上的万米赛跑，运动员要在 400 米的圆形跑道上跑 25 圈.于是当一个运动员一路领先已到了冲刺阶段，金牌在望时，突然看见前面还有一个运动员，他当然可以“置之不理”，因为“前面”的运动员还只跑了 24 圈，还差 400 米呢.上面我们写了一些“游戏文章”其实例如道路、粘结，旋转楼梯等等都大有数学文章在后面.

如果读者仍然感到 $S_n$ 神秘莫测，我们再换一个角度来看：$w=z^{\frac{1}{n}}$ 是 $w=z^n$ 的反函数，因此其定义域是后者的值域.$w=z^n$ 映 $C$：($z$ 平面)$\to S_n$，我们现在研究 $S_n$ 的构造.把 $z$ 平面 $C$ 分成 $n$ 个扇形，其顶角均为 $2\pi/n$，于是 $\overline{OA_0}$ 被映为 $w$ 平面的正实轴，$\overline{OA_1}$ 也是一样，总之第一个扇形被一对一地映为 $\omega$ 的全平面，而 $\overline{OA_0}$ 我们认为变为 $w$ 平面正实轴“上岸”，$\overline{OA_1}$ 变为 $w$ 平面正实轴“下岸”，总之，我们得到第一页的平面.再往下进行到了第二个扇形，$\overline{OA_1}$ 变成第二页 $w$ 平面正实轴的上岸，而与第一页相邻.仿此进行下去可以作出 $n$ 页 $w$ 平面而且再回到第一页.在图 2-3-4 上我们画了一些小区域，它们都是跨在两页 $w$ 平面上的，借此标志第一页与第二页以及第 $n$ 页相邻，但与第三页、第 $n-1$ 页都不相邻.所以在第 $n$ 页上的 $w$ 又会“从天而降”，不知怎么又回到了第一页.这样看来 $w=z^n$ 的值域 $S_n$ 显然具有上面讲的黎曼曲面的构造.所以它的反函数 $w=z^{\frac{1}{n}}$ 应以这样的 $S_n$ 为定义域，而且成为 $S_n$ 上的单值函数.所以，黎曼曲面是实实在在的结构而非空中楼阁！

有了对 $w=\sqrt[n]{z}$ 的详细讨论，再看 $w=\mathrm{e}^z$ 的反函数 $w=\ln z$ 就容易多了.事实上，因为 $z=r\mathrm{e}^{t\varphi}=r\mathrm{e}^{\mathrm{i}(\varphi+2k\pi)}$，$k=0,\pm1,\pm2,\cdots$，

$$w=\ln z=\ln r+\mathrm{i}(\varphi+2k\pi),\quad k=0,\pm1,\pm2,\cdots.\tag{6}$$

从表面上看来，$\ln z$ 是一个无穷值的多值函数，但是如果按上面讨论 $w=\sqrt[n]{z}$ 的方法来处理，可以知道它只是一个定义在黎曼曲面 $S_\infty$ 上的函数.$S_\infty$ 由无穷多页 $z$ 平面叠合而成(图 2-3-5).只不过，它一方面可以向上(如果 $z$ 绕 $z_0$ 作反时针方向旋转的话)升到无穷多层，也可以向下(如果 $z$ 绕 $z_0$ 作顺时针方向旋转的话)，降到无穷多层的“地下室”去.读者由此可见，$w=\ln z$ 是黎曼曲面 $S_\infty$(它有无穷多页)上的单值函数.

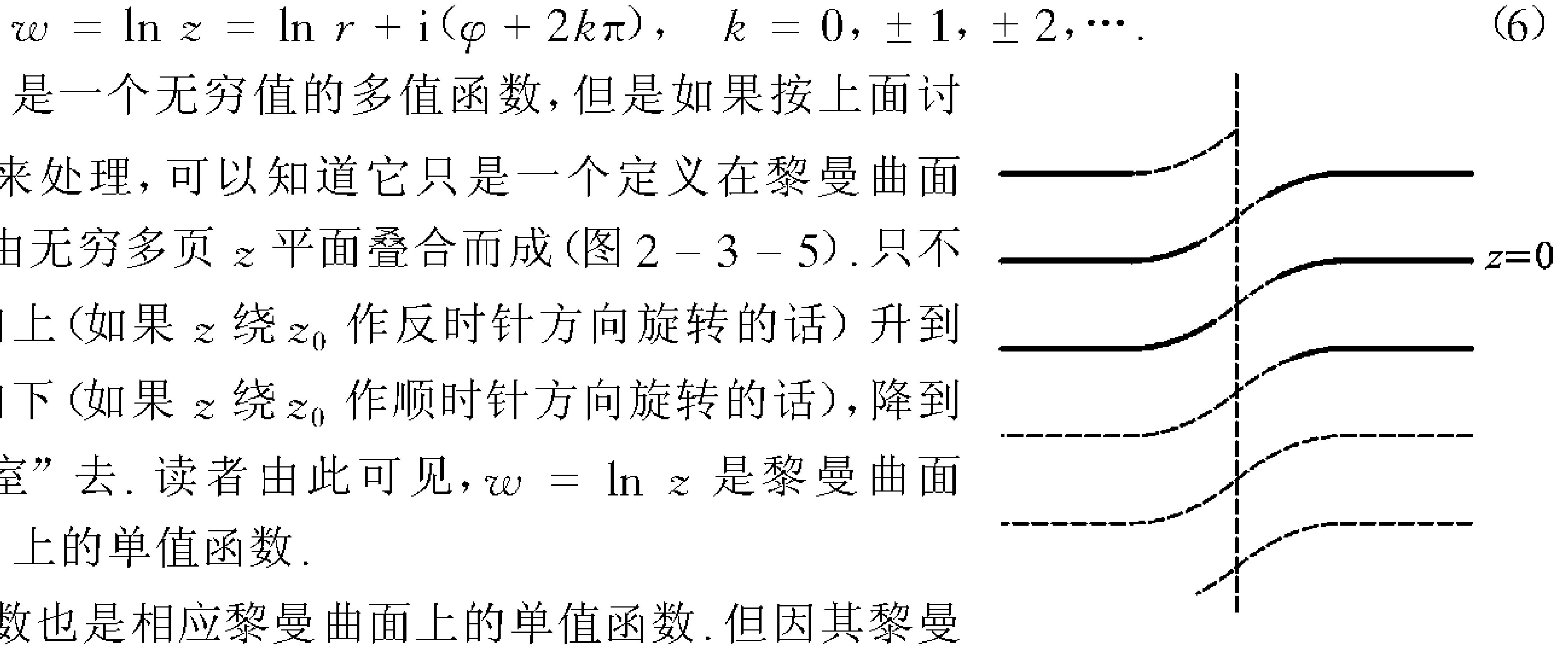



图 2-3-5

还有反三角函数也是相应黎曼曲面上的单值函数.但因其黎曼曲面的构造比较复杂，我们就不再讨论了.

可是不论怎么说，神秘气氛终于难以尽除.为什么数学家们肯

接受这样的东西？根本之点在于它能解决深刻的数学问题和物理问题．从上面讲的处理多值函数的情况看来，一旦进入复域，许多隐藏在水下的东西就浮现出来了．所以，数学家一直努力把它涉及的基本思想弄清楚．到 1913 年外尔（H. Weyl）的名著《黎曼曲面的概念》出版，明确指出，黎曼曲面是一个微分流形，而且对于什么是微分流形也给出了今天我们采用的说法．后来的研究表明了，黎曼曲面不能在 $\mathbf{R}^3$ 中“实现”，而只有在更高维的 $\mathbf{R}^n$ 中才能使它不自交地实现．关于微分流形我们将在第七章中讨论，那时将再回到黎曼曲面的问题．我们现在想要指出的只有一点：函数的定义之内涵不是固定不变的，而是随着数学与物理学的发展，越来越丰富，也越来越深刻地表现了大自然的奥秘．

## §4 “函数”概念够用了吗？

我们先回顾一下函数概念的发展．本章中我们并没有就函数的定义作很多文章，而是抓住了几个函数（其实只有一个函数即指数函数），尽可能地讲到它的方方面面．当然应指出，我们还只涉及一些浅层次的问题，而对于它所蕴含着的更深刻的问题——对称性并未接触．整个第二章实际上都是讲的指数函数．我们这样做的目的在于指明，数学的发展根本的主题是揭示大自然的深层的规律性．各种概念、方法、理论的提出与发展，归根结蒂是服务于此目的的．函数概念当然也不例外．

尽管微积分学研究的主题是函数．微积分的创始人如牛顿、莱尼布茨都只是讨论具体的函数，如幂函数、指数与对数函数、三角函数和一些很特殊的与物理问题相关的曲线如旋轮线、悬链线等等．因为他们研究的对象基本上就是这些．大约到 18 世纪，才开始有了一般的函数概念．例如欧拉就给出了两种“说法”（恐怕不宜说是定义），其一见于他的《无穷小分析引论》（1748），是说函数即变量和常数组合而成的表达式，其中他概括了某些代数函数和某些超越函数，并且看出了它们的区别．可是欧拉还有第二种说法，就说 $y=f(x)$ 即在 $xy$ 平面上“随手画出的曲线”．欧拉的第二个定义的来源显然受到当时关于弦振动问题研究的影响，因为一根弦的形状显然是可以顺手画出来的．在函数概念进一步的发展中，傅里叶的研究起了极大的作用．例如，他研究了单位圆中温度的分布，并假定在边界上温度的分布是这样的：在左半圆周上温度为 1，右半圆周上温度为 $-1$．若用极坐标 $(r,\theta)$ 表示，则若圆盘内各点的温度是 $T(r,\theta)$，则

$$T(1,\theta)=f(\theta)=\begin{cases}1, & |\theta|\leqslant\dfrac{\pi}{2},\\ -1, & |\theta|\geqslant\dfrac{\pi}{2}.\end{cases}$$

这里 $f(\theta)$ 用两个表达式来表示，至今仍有不少人把它称为“分段函数”，并且花不少精力来讨论它．其实它就是一个函数，是 $T(r,\theta)$ 在 $r=1$ 处的边值．傅里叶给出了这个函数的展开式如下：

$$\frac{4}{\pi}f(\theta)=\cos\theta-\frac{1}{3}\cos 3\theta+\frac{1}{5}\cos 5\theta-\cdots$$

因为这个例子常被人引用，所以很有点名气，本书中就引用了好几次．其实这种具有间断点的函数的傅里叶展开式例子很多，也都容易算，上面的例子并没有什么“特色”．许多人都以为“完全任意的函数”（这是傅里叶的原话）都可以展开为收敛的傅里叶级数，傅里叶也是其中之一．狄利克

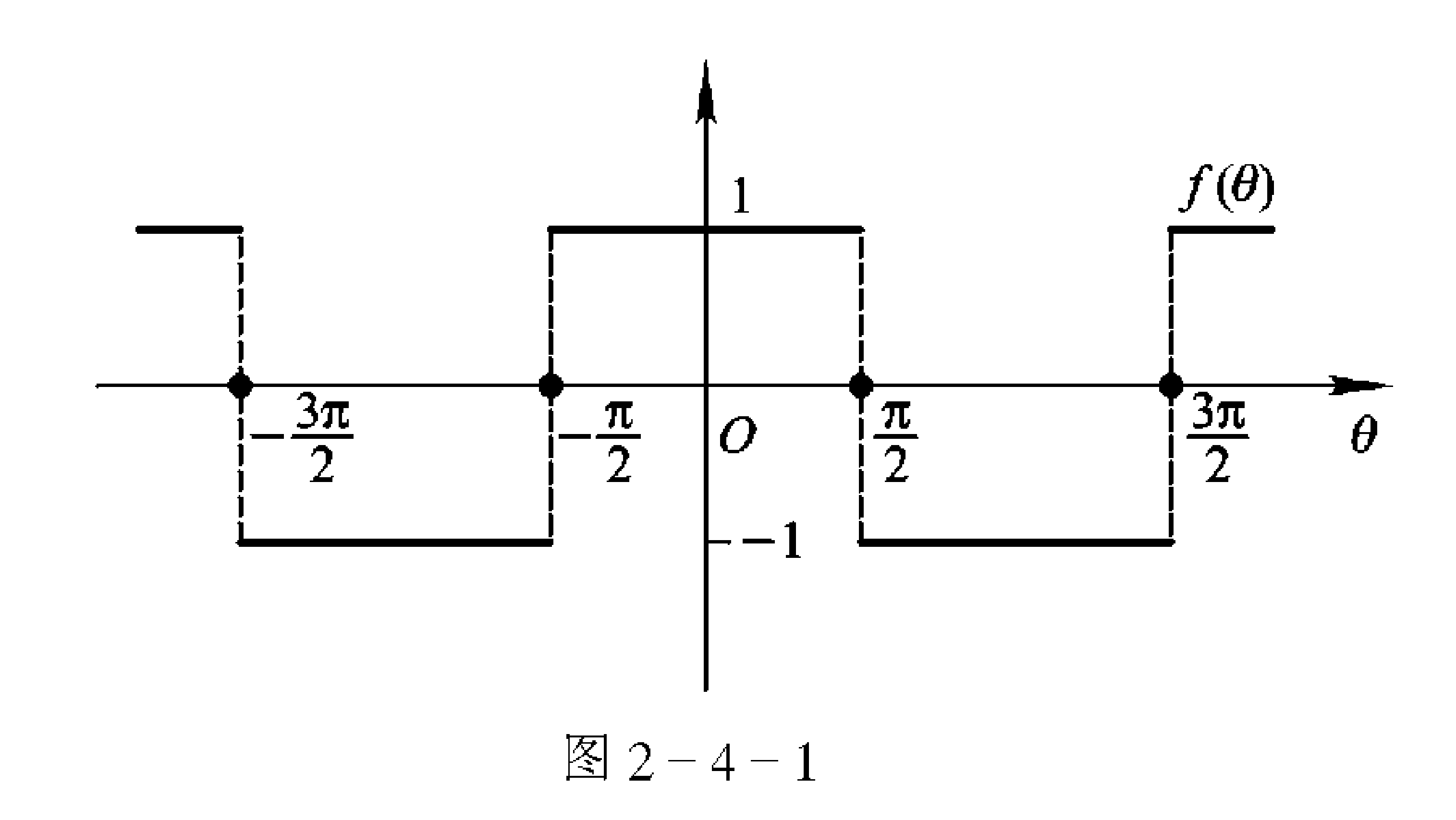



图 2-4-1

雷则发现了,在满足一组相当复杂的条件(现在常称为狄利克雷条件)后,函数 $f(x)$之傅里叶级数才确实收敛.可是若 $x_0$是 $f(x)$的连续点,则级数在 $x_0$之和就是 $f(x_0)$.若 $x_0$是 $f(x)$的第一类间断点,则级数在 $x_0$之和却是$\frac{1}{2}(f(x_0+0)+f(x_0-0))$.对上面的例子,它在 $x=-\frac{3}{2}\pi$, $-\frac{\pi}{2},\frac{\pi}{2},\frac{3\pi}{2}\cdots$ $\left(总之在 \pm\frac{2k+1}{2}\pi\right)$处,收敛于图 2-4-1 上的小圆点.所以这个级数之和并不是前面的式子$\left(还有因子\frac{4}{\pi}\right)$,而如果要写准确,在$[-\pi,\pi]$内应是

$$f(\theta)=\begin{cases}1, & |\theta|<\frac{\pi}{2},\\ 0, & |\theta|=\frac{\pi}{2},\\ -1, & \frac{\pi}{2}<|\theta|\leqslant\pi.\end{cases}$$

(细心的读者会发现前式中我们既然写了 $|\theta|\leqslant\frac{\pi}{2}$,所以在 $|\theta|=\frac{\pi}{2}$ 时两个式子矛盾了.我们这里是"故意"这样做的,因为今后我们会看到,若在"个别点"上改变函数值,有时并不影响讨论,详见第五章.但在目前的讨论中,确实应写为后式).那么上式是不是仍表示一个函数?它既不是一个解析表达式,又不是可以"顺手"画出的曲线.这些事实迫使狄利克雷给出定义在区间 $A$ 上的函数的定义,即"有两个变量 $x$ 和$y$,$x\in A$,而且对每个 $x$ 都有一个$y$ 与之相应",请读者注意,我们在定义中没有说什么"有确定的规律"或"有一定的方法"找出与 $x$ 对应的$y$.因为什么叫"规律",什么叫"方法"都是说不清的.是不是指某种"算法"?确有这样的函数,它是没有一定的算法来计算的,如此等等.这当然都是后来的人提出的问题(但决非无中生有的故作高深),狄利克雷本人不仅没有想到这一步,而且有时仍然说"完全任意的函数"都可以展开为傅里叶级数.其实他心里想的是适合狄利克雷条件的函数!由此可以想见传统的束缚和习惯的因素是多么有力!

傅里叶级数确实是一个产生重大数学成果的来源.狄利克雷已考虑了具有第一类间断点的函数,如果有无穷多个间断点,那么可以考虑间断点的极限点.如果极限点只有有限多个还好说,有无穷多个又怎么办?是不是还应该考虑极限点的极限点呢?这样一来,人们所研究的函数范围就扩大到前人无法思议的地步.这时把函数定义为解析表达式或者某种图像都是无所助益的了.可是为什么要研究这种"病态"的函数?我们在第三章§7 中将讲一些理由,而整个第四章实

际上是围绕着这个问题来展开的：函数的类型越来越复杂，其反映大自然规律的深度越来越深．而到第六章，则对这些极丰富而又多样的对象给出了相当统一的框架，而丝毫不失严格性．到那里再来看初学者视为梦魇的 $\varepsilon-\delta$，就会感到还不过是雕虫小技，而有“会当凌绝顶，一览众山小”之慨！

那么，函数概念发展到这一步就“够用”了吗？问题的核心是某些自然现象所表现出的奇异性是古典的微积分（亦即所谓的古典分析）难以应对的．这里的情况和微积分的发展有些类似：人们在很长的时间里只能用一种不严格的、直观的方法，或者说是在没有充分根据的情况下沿用古典分析的概念与方法，然而总是得到了正确的结果．到了 20 世纪 30 年代，在许多数学物理问题中都出现了这种情况，使人们不能不认真地对待它们．然而幸运的是，到了 20 世纪 50 年代就系统地解决了这个矛盾，出现了所谓广义函数论（这是前苏联数学家的称呼，而西方数学界则称为分布理论）．我们在这里不可能详细地介绍它，而要在四、五两章中讲解其若干基本点．现在我们从一个最简单、而且谁都认为不可回避的问题开始．我们的讲法大体上相应于 20 世纪中叶没有掌握广义函数论的物理学家处理这类问题的方法，但随时指出其不足之处．但在本节最后，我们却要讲一个十分重要的定理．我们对它的处理将是严格的，而且证明的思想却很好地表现了广义函数的力量．

假设有一根棒，我们不妨设它是整个实数轴．如果从最左端算起直到 $x$ 点的这一段，其质量是 $m(x)$（当然我们不妨设 $m(x)$ 是有限的，因为总质量为 $\infty$ 的棒是很难设想的）．我们来算一下，它在各点处的密度．为此，在棒上取 $(x, x+\Delta x)$ 一小段，其质量应该是 $\Delta m = m(x+\Delta x) - m(x)$，所以其平均密度是 $\dfrac{\Delta m}{\Delta x} = \dfrac{m(x+\Delta x) - m(x)}{\Delta x}$，而 $x$ 点的密度是一个函数 $\rho(x) = \dfrac{\mathrm{d}m(x)}{\mathrm{d}x}$．只要 $m(x)$ 可求导，以上的作法就是合理的．又因 $m(x)$ 是从最左端 $x=-\infty$ 算起的，所以 $m(-\infty)=0$，因此又有 $m(x) = \int_{-\infty}^{x} \rho(x)\mathrm{d}x$，只要 $\rho(x)$ 是比较“规则”的，例如是连续的，而且这个积分又是收敛的，这个作法也是合理的．总之，在适当的“光滑性”（光滑一词就是指直到某阶导数均为连续）以及 $\rho(x)$ 的可积性条件下总有

$$\rho(x) = \frac{\mathrm{d}m(x)}{\mathrm{d}x},$$

$$m(x) = \int_{-\infty}^{x} \rho(x)\mathrm{d}x. \tag{1}$$

可是我们来看，若有一个单位质量的质点放在 $x=0$ 处，而棒的其它地方都没有质量，这时还有没有密度？(1) 式是否成立？这个情况当然也是理想化的，但是是很自然的．先计算 $m(x)$．若 $x<0$，则 $m(x)=0$，若 $x\geqslant 0$，则 $m(x)=1$，所以 $m(x)$ 就是赫维赛德（O. Heaviside，1850—1925）函数：

$$m(x) = H(x) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0. \end{cases} \tag{2}$$

读者会看到，若限于 $-\pi<\theta<\pi$，则上面我们讲的 $T(1,\theta) = f(\theta) = 2\left[H(\theta) - \dfrac{1}{2}\right]$．在那里我们说，在 $|\theta| = \dfrac{\pi}{2}$ 处，$f(\theta)$ 用哪个式子来表示并不重要．但现在却严格规定了 $H(0)=1$．这是因为在

$x=0$ 处有一个单位质量存在.赫维赛德是一个电气工程师,他研究一个电路.如果在某一瞬间 $x=0$(这里 $x$ 表示时间)把电路接通,输入一个单位电流,则在 $x\geqslant 0$ 时,电流为 1,$x<0$ 时,电流为 0,他认为这种情况在研究电路理论中是最基本的.在某种意义下,这就是广义函数论的开始,时为 1890 年左右.

这种质量分布的密度是什么呢?现在我们暂把数学严格性的要求置诸脑后.如果 $x<0$,则只要 $\Delta x$ 充分小,仍有 $x+\Delta x<0$,因此

$$\Delta m = m(x+\Delta x) - m(x) = 0 - 0 = 0,$$

$\frac{\Delta m}{\Delta x}=0$,其极限也为 0.即当 $x<0$ 时,$\rho(x)=0$.若 $x>0$,则当 $\Delta x$ 充分小时,仍有 $x+\Delta x>0$,所以 $\Delta m=m(x+\Delta x)-m(x)=1-1=0$,我们仍有 $\rho(x)=0$.可是,注意,这里开始出问题了,如果我们把 $x$ 放在 $x=0$ 的"紧左边",从而 $m(x)=0$,而若 $x+\Delta x$ 在 $x=0$ 的"紧右边",则 $m(x+\Delta x)=1$,从而$\frac{\Delta m}{\Delta x}=1/\Delta x$,而

$$\rho(0) = \infty.$$

大家立刻会看到,这已经不是数学语言了:什么叫紧左边,什么叫紧右边?无非是要把这个质点放在区间$[x,x+\Delta x]$内,我们取极限时不只令 $\Delta x\to 0$,而且也令 $x\to 0$.这当然好办,我们只需取区间 $\left[-\frac{\Delta x}{2},\frac{\Delta x}{2}\right]$,$\Delta x>0$ 即可.可是这个极限是不存在的.习惯了数学思维的读者会问,怎么能够把不存在的东西叫密度呢?可是对物理学家,包括狄拉克(P. A. M. Dirac,量子物理学的大师)在内,却认为这是非常自然的.试想区间之长 $\Delta x$ 是一个无穷小,可是这个区间中的质量 $\Delta m=1$,二者一除不是无穷大又是什么呢?狄拉克当时的想法也就在这个水平上.如果伯克莱主教再世,他很可能会写一篇《不信神的物理学家》,并且说:狄拉克先生,您连量子力学那样希奇古怪的东西都相信,凭什么不相信上帝呢?带着罪恶的灵魂怎么能进天国呢?难怪哲学家拉卡托斯(J. Lakatos)在《证明与反驳》一书中要再调侃一番了:牛顿等了两百多年,才有人出来拯救他的灵魂,进了数学的天堂,狄拉克要幸运得多,他还在世时,施瓦兹(L. Schwartz,有两个施瓦兹,这一位是法国数学家,另一位是 H. A. Schwarz,德国人,魏尔斯特拉斯的高足,我们每日每时都在用他的不等式.你不要忘了,我们每时每分都用的 $\varepsilon-\delta$ 并不是发表在什么 SCI 刊物上,而是见于他听自己的老师讲课所记的笔记中.为了区别,以后都用施瓦茨来表示德国人;而法国人都译为"施瓦兹".——本书作者注.)就拯救了他的灵魂.狄拉克于 1984 年去世,可是施瓦兹在 1950 年就发表了《分布理论》(这个名词的来源很清楚了,我们讨论的不是一种质量分布吗?)一书,完全符合现代数学严格性要求,处理了狄拉克的理论.

这里有一个重要的争论:物理学家会说,难道狄拉克需要人来拯救他的灵魂吗?他已经在天国安息,当然天国里也会给施瓦兹留下位置.可是狄拉克一定会坐在更加紧靠上帝的椅子上.施瓦兹的"灵魂"的一部分,即施瓦兹核定理,不是狄拉克早已知道,并以自己的语言来陈述过?这可是你们数学家说的,请看冯•诺依曼(Von Neumann)《量子力学的数学基础》25~27 页.数学家当然也可以反唇相讥:不妨再看一下爱因斯坦,难道不是黎曼给了爱因斯坦以灵魂吗?这可是爱因斯坦自己也承认的.这样的争论已经有多少年了,还会继续多少年?看来,数学与物理学是两种不同的文化,各有自己的方法与风格,然而都有共同的目标——探讨大自然的奥秘,而且携手并肩,结伴前进.

不论如何，我们得到了一个“函数”，狄拉克称之为“$\delta$ 函数”而且现在有公认的记号 $\delta(x)$：

$$\delta(x)=\begin{cases}\infty, & x=0,\\ 0, & x\neq 0.\end{cases} \tag{3}$$

但读者可以看出一个问题：如果我们在 $x=0$ 处放一个质量为 2 的质点，则一方面看，我们应得到 $2\delta(x)$，另一方面，我们又看到，会得到 $\dfrac{\Delta m}{\Delta x}=\dfrac{2}{\Delta x}$，仍是无穷大，归根结蒂：$2\infty=\infty$，因此仍会得到 $\delta(x)$. 为了区别这些情况，我们在定义 $\delta$ 函数时还需要附加一个表示总质量为 1 的条件：

$$\int_{-\infty}^{\infty}\delta(x)\mathrm{d}x=1. \tag{4}$$

$\delta$ 函数即由(3)，(4)两式来“定义”. 我们现在可以把这个“定义”中“不合法”的地方再重复一下：首先，一个函数应该对 $x$ 之每个值均有一个确定值与之对应，而 $\infty$ 肯定不能算是一个确定的值. 其次，一个函数在某些点没有定义也是常有的事，这时定义其积分时常也是可以的，但按(3)式，$\delta(x)$的“积分”应该为 0，而不应该是(4)式.

$\delta$ 函数最早是由赫维赛德在 1894—1895 年间定义的. 1927 年狄拉克在他关于量子力学的基本论文(1927 年)提出以上的讲法，可以在狄拉克《量子力学原理》第三章 §15 (P. A. M. Dirac, Principles of Quantum Mechanics, 4th ed., Clarendon Press, 1958, 58～61 页)中找到. 在他的基本论文(The physical interpretation of the quantum dymamics, Proc. of the Royal Soc. of London, A113, (1927)621～641)中，他说“当然，严格说来，$\delta(x)$并不是 $x$ 在本来意义下的函数，但是可以看作是某个函数序列的极限. 完全同样，把它用于量子力学，对其几乎所有的需要，都不会得出不正确的结果”.

狄拉克在那本书和那篇论文中提出了两种解释 $\delta$ 函数的方法，一是利用对偶性(这是现代的语言，狄拉克那时是说用分部积分法)，一是利用函数序列的极限. 下面我们再加上一种利用测度的方法，则来自施瓦兹，其实也是对偶性之一例. 我们将在第四章中讲解对偶性与函数序列的“极限”，现在我们先来不太严格地介绍测度方法，使读者有机会接触一点现代数学的思想和方法.

这本书里将不会一般地讨论测度理论. 我们只直观地指出，测度是一个“集函数”. 以一维情况为例，如果作一个分划

$$-\infty<\cdots<a_1<\cdots<a_{n-1}<a_n\cdots<+\infty,$$

则可得一系列左闭右开的区间 $\sigma_n=[a_{n-1},a_n)$把实数轴 $\mathbf{R}$ 覆盖起来：$\mathbf{R}=\bigcup\limits_{n=-\infty}^{+\infty}\sigma_n$，对每个区间 $\sigma_n$，我们规定一个实数 $m(\sigma_n)$，使得

$$m(\bigcup\sigma_n)=\sum m(\sigma_n). \tag{5}$$

这样的 $m$ 当然是一个“集函数”，即对集 $\sigma_n$ 确定了一个实数 $m(\sigma_n)$ 与之相应. 测度的例子很多，通常的区间长度是一个测度，如上面所说的棒上的质量分布也是测度. 这里有两点需要说明：首先我们是从区间 $\sigma_n$ 而不是更一般的集合开始，而且(5)式中的 $\bigcup$ 与和 $\sum$ 都不明确规定是有限并(有限和)还是可数并(可数和). 我们在这里虽然含糊其辞了，好在读者们在第四章会有机会读到准确的论述. 其次，我们取 $\sigma_n$ 为左闭右开的而且彼此不相交，这却是重要的. 因为我们要考虑集中在某一点的质量，如果该点在 $x=a_n$，则这个点的质量是算在$[a_{n-1},a_n]$的质量内还是算

在$[a_n, a_{n+1}]$内呢?我们不能把一个质点的质量算两次,那会使棒的总质量产生疑义.总之,我们称适合(5)的$m(\sigma)$叫做测度.有了测度就可以定义积分:设$f(x)$充分光滑,而且在某有限区间$[A,B]$之外恒为0,则可以定义$f(x)$关于测度$m(\sigma)$的积分为

$$\int f(x)\mathrm{d}m = \lim\sum f(\xi_i)m(\sigma_i). \tag{6}$$

这个积分称为黎曼—斯蒂尔切斯(Stieltjes)积分,它与黎曼积分的区别在于黎曼积分使用了一个特别简单的测度:$m(\sigma_n)=(a_n-a_{n-1})$.$\delta$函数则相应于上面讲的有一个单位质量的质点位于$x=0$处的测度.我们很容易计算这时的(6)式,设$x_0=0$位于$[a_j, a_{j+1})$内,则(6)式右方各项除了$i=j$一项外全为0,从而只余下一项$f(\xi_j)\cdot 1$,$\xi_j\in[a_j, a_{j+1})$.由于我们假设了$f(x)$充分光滑,故有

$$\int f(x)\mathrm{d}m = f(0). \tag{7}$$

这里我们必须在观念上有一个大转变:过去讲函数总是讲的可以在某一集合上逐点取值的"东西".(当然"逐点取值"有时可以马虎一点,在积分一章中我们还要回来讲怎样"马虎").现在,$\delta$函数则对应于一个测度,测度可不是逐点取值的东西,而是在某个集合$\sigma_n$上有一个值.但是,它可以作用在一个充分光滑的函数$f(x)$上,按(7)式起一种作用,即取出该函数(以后称为试验函数)在$x=0$处的值来.凡是可以按某种线性的规律如(7)在一类试验函数上取值的对象,叫做一个线性泛函.一个普通的连续函数$\varphi(x)$也可以在一类试验函数上取值,例如按下式

$$(\varphi, f) = \int \varphi(x)f(x)\mathrm{d}x \tag{8}$$

取值.如果$\varphi(x)$有原函数$\Phi(x)$,则(8)式还可以写为

$$(\varphi, f) = \int f(x)\mathrm{d}\Phi.$$

所以,普通的连续函数也可以形成泛函.但是泛函不一定由普通连续函数生成,而也可以由例如测度生成.所以我们把这些线性泛函叫做广义函数.其实还可以由更加复杂的东西生成更复杂的广义函数,$\delta$函数只是最简单的一例,它按

$$(\delta, f) = f(0) \tag{9}$$

生成一个由测度给出的广义函数,就是上文讲的在$x=0$处有单位质量的质点生成的测度.

这样,我们看到测度可以认为是函数概念的推广.测度理论是一个很重要的数学分支,它特别是概率论的基础.我们将在第四章讨论一个特别有用的测度——勒贝格测度以及勒贝格积分.它是黎曼积分直接的推广.至于更一般的测度和积分(包括斯蒂尔切斯积分)我们都不讲了.我们只指出,它们都构成广义函数,而未在广义函数理论中把这一类划分出来专门讨论,因为那是太费力了.

再看如何用分部积分法来解释$\delta$函数.和上面一样,取$f(x)$为一试验函数,即定义在$(-\infty,+\infty)$上的充分光滑而且在一有限区间$[A,B]$外恒为0的函数.若$\varphi(x)$在$(-\infty,\infty)$上有连续的一阶导数,则有

$$\int_{-\infty}^{\infty}\varphi(x)f'(x)\mathrm{d}x = \int_A^B\varphi(x)f'(x)\mathrm{d}x = \varphi(x)f(x)\Big|_A^B - \int_A^B\varphi'(x)f(x)\mathrm{d}x$$

$$= -\int_A^B \varphi'(x) f(x) \mathrm{d}x = -(\varphi', f). \tag{10}$$

这里我们利用了试验函数 $f(x)$ 充分光滑且在 $[A,B]$ 外恒为0，从而 $f(A) = f(B) = 0$. (10) 告诉我们，若 $\varphi'(x)$ 连续，则 $(\varphi, f')$ 可以化为 $\varphi'$ 作用在 $f$ 上形成的泛函（再改变符号），但是可能有这样的情况，即 $\varphi'(x)$ 不一定是一个好的函数，但 $(\varphi, f')$ 仍可以化为作用在 $f(x)$ 上所成的泛函（再改变符号）. 这时，何妨把这个泛函看作是 $\varphi(x)$ 在某种广义意义下的导数？这是数学中推广某个概念的常用的方法：某个对象（如这 $\varphi(x)$）在某种条件下（这里是：$\varphi'(x)$ 连续）具有某种性质（这里是：在试验函数 $f(x)$ 上产生一个泛函如 (8) 式），而另一个对象在缺少该条件时也具有此种性质，我们就说这个另外的对象在缺少该条件时，广义地具有此种性质. 我们来看一下“另一个对象”的例子：即赫维赛德函数 (2). 它不是连续可微的，因为 $H(x)$ 在 $x = 0$ 时甚至不连续，但若按 (8) 式将它作用到试验函数 $f(x)$ 之导数 $f'(x)$ 上，我们有

$$\begin{aligned}\int_{-\infty}^{\infty} H(x) f'(x) \mathrm{d}x &= \int_0^B H(x) f'(x) \mathrm{d}x \\ &= \int_0^B f'(x) \mathrm{d}x = f(x)\Big|_0^B = -f(0) = -(\delta, f).\end{aligned}$$

右方除了相差一个符号外即是 (9) 式，因此，我们说 $\delta$ 函数即 $H(x)$ 在广义函数意义下的导数：

$$H'(x) = \delta(x).$$

这就完全解释了赫维赛德的思想的合理内核.

我们再来回顾一下这里的作法. 我们有两种对象，其一是试验函数 $f(x)$，它的集合是一个线性空间. 而其元素都是性质很好的函数：我们这里规定 $f(x)$ 充分光滑，而且在某有限区间之外为 0. 另一类对象如 $\varphi(x)$，也形成了一个线性空间，例如对于 $\varphi_1(x)$，$\varphi_2(x)$ 我们也可以研究 $c_1\varphi_1(x) + c_2\varphi_2(x)$ 等等. 它的元素则可以不太“规矩”，例如赫维赛德函数在 $x = 0$ 不连续. 这些对象甚至可以是测度，根本不是逐点取值的函数. 但是，它们可以像连续函数一样，在试验空间的函数上生成线性泛函. 只不过生成的方式可以是 (9)，而不一定是 (8). 这个空间称为试验函数的对偶空间. 这个名词也是由线性代数中借来的，而例如 $\int \varphi(x) f(x) \mathrm{d}x = (\varphi, f)$ 就好像线性代数中两个向量的内积一样. $\varphi(x)$ 虽然不规则，例如不能求导，但是 $f(x)$ 则很规则可以求导，而且我们可以利用 (10) 式，把对 $\varphi(x)$ 求导这个困难问题转移到对 $f(x)$ 求导上去. 这就是对偶性的妙用. 分部积分法就可以看成是利用对偶性把“困难”由 $\varphi(x)$ 身上“转嫁”到 $f(x)$ 身上的办法. 因此是一个很深刻的公式.

广义函数就是试验函数空间上的线性泛函，它是线性代数的思想与微积分的思想融合的产物. 它使我们摆脱了函数必须在某一集合之各点上取值这个限制，使得许多分析运算可以很顺利地推广到原来不能应用的领域，帮助我们研究物理科学中许多有高度奇异性的对象.

下面我们再来解释狄拉克“$\delta(x)$ 可以看作是某个函数序列的极限”这句话. 狄拉克是一个伟大的物理学家. 他高度尊崇数学对物理学的重要性. 他甚至认为数学美是选择在理论物理学中走哪一条路的最终判据. 他说“方程之美比之其与实验结果的符合更重要”. 当二者不甚符合时“不必太丧气，因为二者不符可能是由于某些较小的地方……而在理论的进步中迟早会弄清楚的.”但是他又不甚注意数学的严格性. 他本是在英国布里斯托 (Bristol) 大学学电工的. 后来 (1923) 才到剑桥圣约翰学院做研究生，转入了理论物理领域. 1932 年他当选为剑桥大学卢卡斯

数学讲座教授(Lucasian Professor of Mathematics),这是当年牛顿担任过的教职.狄拉克在大学中学的主要是工科数学.后来,他这样回顾自己的经历:“我愿解释一下工程师的训练和工程教育对于我所起的作用……过去,我只对完全准确的方程有兴趣.然而我受到的工程训练教会了我要容忍近似.我能够看到有时近似的东西也包含了相当的美.……我想,如果不是有过这种工程训练,我后来作的那一类工作是不会成功的.……我在后来的工作中主要是继续使用了工程师的非严格的数学.我想,你们会发现我后来的著作中都用了非严格的数学”.在狄拉克看来,一是对数学美的追求,二是卓越的洞察力才是成功的保证.这是非常值得我们深思的.下面我们将先用完全严格的方法证明一个著名定理,再从 $\delta$ 函数的观点去重新看待它,作为对狄拉克的方法论的一点小小回应,并就此结束我们对函数概念的讨论.

**魏尔斯特拉斯定理**,设 $f(x)$ 为某有限闭区间上的连续函数.则必可找到一串多项式 $P_n(x)$ 在此闭区间上一致收敛于 $f(x)$.

**证**　设 $f(x)$ 是 $[\alpha,\beta]$ 上的连续函数,这里 $0<\alpha<\beta<1$.可以把 $f(x)$ 连续地拓展到 $(-\infty,\infty)$ 上,而且使它在 $[-A,A](A>1)$ 外恒为 0,这样一来 $f(x)$ 就是一个试验函数了(试验函数的“充分光滑性”本来就有相当的活动余地,这里我们只要求它是连续的.这一段话只与以下讨论 $\delta$ 函数有关,与本定理之证明无关).

考虑积分

$$I_n=\int_{-1}^{1}(1-u^2)^n\mathrm{d}u,\tag{11}$$

将它分成两部分,一部分是在 $|u|\leqslant n^{-1/3}$ 上积分,另一部分是在 $n^{-1/3}\leqslant|u|\leqslant 1$ 上积分,即

$$I_n=K_n+L_n=\int_{-n^{-1/3}}^{n^{-1/3}}+\int_{n^{-1/3}\leqslant|u|\leqslant 1}.$$

当 $n\to\infty$ 时,$K_n$ 之积分区域缩为一点,但是它的贡献却是几乎 $I_n$ 的全部:

$$\lim_{n\to\infty}L_n/I_n=0.$$

证明如下:在 $L_n$ 中,被积函数之最大值在 $|u|=n^{-1/3}$ 时达到,所以

$$L_n<2(1-n^{-2/3})^n.$$

$I_n$ 之积分区域则包含了 $\left[-\frac{1}{2}n^{-1/3},\frac{1}{2}n^{-1/3}\right]$,所以

$$I_n>\int_{\frac{1}{2}n^{-1/3}}^{\frac{1}{2}n^{-1/3}}(1-u^2)^n\mathrm{d}u>\left(1-\frac{1}{4}n^{-2/3}\right)^n\cdot n^{-\frac{1}{3}},$$

因此

$$L_n/I_n<2n^{1/3}\left(\frac{1-n^{-2/3}}{1-\frac{1}{4}n^{-2/3}}\right)^n.$$

但是

$$\frac{1-n^{-2/3}}{1-\frac{1}{4}n^{-2/3}}=1-\frac{\frac{3}{4}n^{-2/3}}{1-\frac{1}{4}n^{-2/3}}\leqslant 1-\frac{3}{4}n^{-2/3}<\mathrm{e}^{-\frac{3}{4}n^{-2/3}},$$

这里我们用到了不等式 $e^x > 1 + x\,(x \neq 0)$，计算一下 $e^x - 1 - x$ 之最小值即可得此式. 总之

$$L_n / I_n < 2n^{1/3} e^{-\frac{3}{4}n^{1/3}} \to 0.$$

现在我们可以来构造 $P_n(x)$ 了，我们需要的 $P_n(x)$ 是一个 $2n$ 次多项式

$$\begin{aligned} P_n(x) &= \frac{1}{I_n}\int_0^1 f(y)\,[1 - (y - x)^2]^n \mathrm{d}y \\ &= \frac{1}{I_n}\int_{-x}^{1-x} f(x + t)\,(1 - t^2)^n \mathrm{d}t. \end{aligned}$$

用上面的式子并用二项定理把 $[1 - (y - x)^2]^n$ 展开，就知道 $P_n(x)$ 是 $x$ 的 $2n$ 次多项式. 变换 $y = x + t$ 则给出第二个式子. 因为 $\alpha \leqslant x \leqslant \beta$，故

$$-x \leqslant -\alpha < 0 < 1 - \beta \leqslant 1 - x.$$

取 $n$ 充分大，使 $n^{-1/3} < \min\{\alpha, 1 - \beta\}$，则

$$-x < -n^{-1/3} < n^{-1/3} < 1 - x.$$

这样，

$$P_n(x) = \frac{1}{I_n}\int_{-x}^{n^{-1/3}} + \frac{1}{I_n}\int_{-n^{-1/3}}^{n^{-1/3}} + \frac{1}{I_n}\int_{n^{-1/3}}^{1-x} = J_1 + J_2 + J_3.$$

$$\begin{aligned} |J_1 + J_3| &= \frac{1}{I_n}\left| \left\{\int_{-x}^{-n^{-1/3}} + \int_{n^{-1/3}}^{1-x}\right\} f(x + t)\,(1 - t^2)^n \mathrm{d}t \right| \\ &\leqslant \frac{M}{I_n}\left\{\int_{-1}^{-n^{-1/3}} + \int_{n^{-1/3}}^{1}\right\}(1 - u^2)^n \mathrm{d}u \\ &= ML_n / I_n \to 0, \qquad M = \max|f(x)|, \end{aligned}$$

这里趋于 0 是对 $x$ 一致趋于 0. 对于 $J_2$ 则利用积分中值定理而有

$$J_2 = \frac{1}{I_n}\int_{-n^{-1/3}}^{n^{-1/3}} f(x + t)\,(1 - t^2)^n \mathrm{d}t = f(x + \theta)\,\frac{K_n}{I_n}, \quad |\theta| \leqslant n^{-1/3},$$

再利用 $f(x)$ 的一致连续性以及 $I_n = L_n + K_n$ 即知对 $x$ 一致地有

$$J_2 \to f(x).$$

定理证毕.

魏尔斯特拉斯定理有许多发展. 例如设 $f(x_1, \cdots, x_n)$ 为连续函数，而且对每个变量 $x_i$ 都有周期 $2\pi$，则 $f(x_1, \cdots, x_n)$ 一定可以用三角多项式

$$\sum a_{k_1, \cdots, k_n} e^{\sqrt{-1}(k_1 x_2 + \cdots + k_n x_n)} \qquad |k_1|, \cdots, |k_n| \leqslant N$$

去一致逼近.

现在我们从 $\delta$ 函数的观点来看看这个证明. 我们很容易想到 $I_n \to 0$，因为计算这个积分并不难. 所以 $\frac{1}{I_n}(1 - u^2)^n$，当 $n$ 越变越大时，图形大抵如图 2-4-2，当 $u \neq 0$ 时，“总有” $\lim\limits_{n\to\infty}\frac{1}{I_n}(1 - u^2)^n = 0$，但当 $u = 0$ 时，$\lim\limits_{n\to\infty}\frac{1}{I_n}(1 - u^2)^n = \infty$，同时恒有 $\frac{1}{I_n}\int_{-1}^{1}(1 - u^2)^n \mathrm{d}u = 1$，所以 $\frac{1}{I_n}(1 - u^2)^n$ 之“极限”就是 $\delta$ 函数，而我们上面证明的定理其实就是说

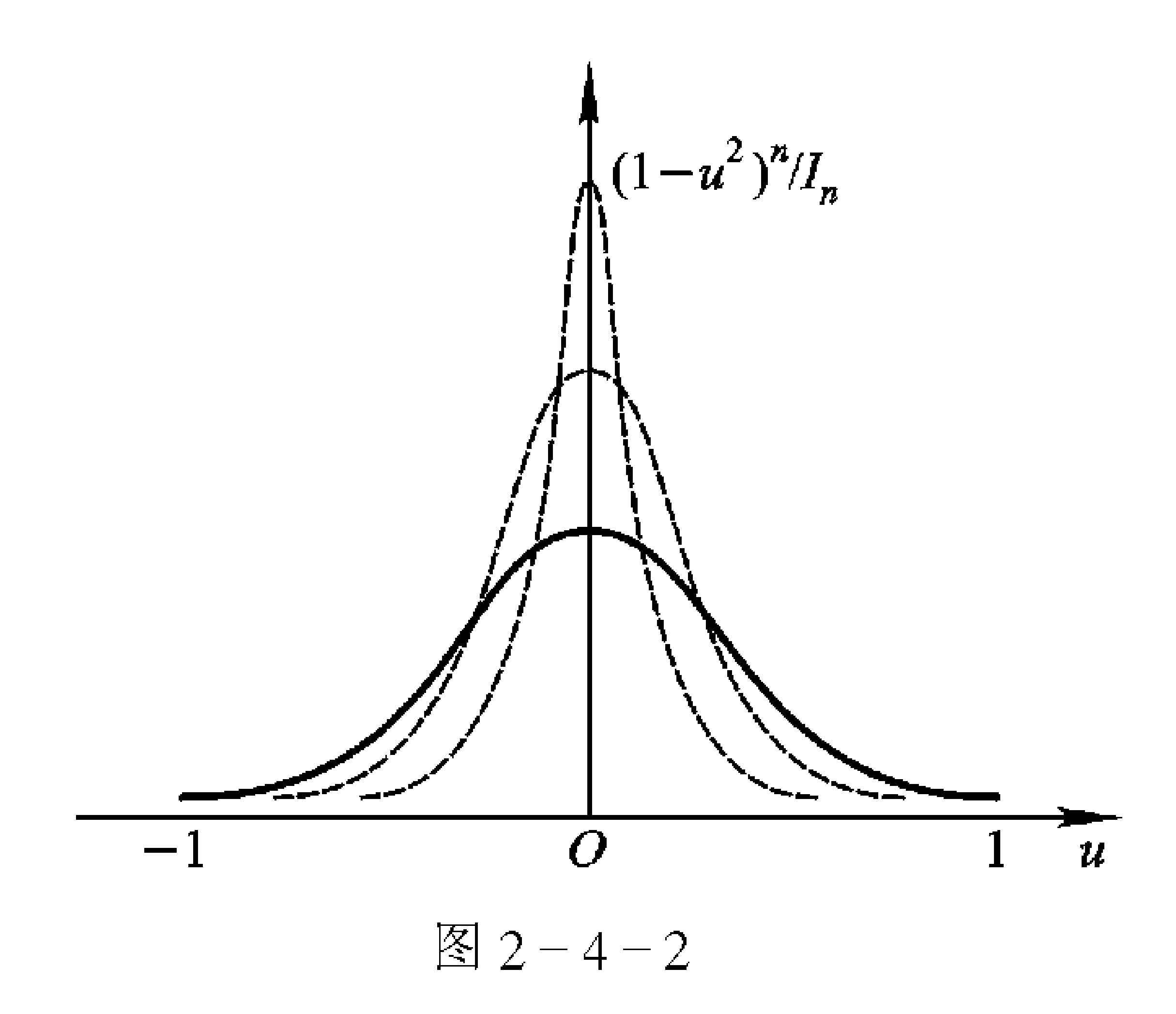



图 2-4-2

$$\lim\int \frac{1}{I_n}(1-t^2)^n f(t+x)\mathrm{d}t = \int \delta(t) f(t+x)\mathrm{d}t = f(x).$$

这就是狄拉克所说的 $\delta(x)$“可以看作是某个函数序列的极限”的意思.但是,上面的讲法总是不严格的.因为只看 $\lim\limits_{n\to\infty}\dfrac{1}{I_n}(1-u^2)^n=\infty\ (u=0$ 时) 以及 $\lim\limits_{n\to\infty}(1-u^2)^n=0\ (u\neq 0$ 时),不能就说函数极限就是 $\delta$ 函数

$$\delta(u)=\begin{cases}\infty, & u=0,\\ 0, & u\neq 0.\end{cases}$$

再则,我们不能随意地在积分号下取极限,所以我们又只好说是一种“广义的”极限过程.我们称它是“弱极限”.但是应说明,“弱极限”是一个很严肃的数学概念,读者以后有很多机会遇见它.上面讲的函数序列,通常称为 $\delta$ 型序列,它在物理学中是很常见的.其中有一些是如上所示的“钟形曲线”,而有一些则完全令人联想不到这种图形,而表现出的剧烈的振荡特性.关于广义函数,包括这里讲的 $\delta$ 型序列,四、五两章中将比较详细地介绍.

# 第三章　微　分　学

## §1　微分学的基本思想

**1. 新的历史条件与新的要求**　我们在第一章中就已讲到，关于运动和变化的考察从古希腊就已开始，而且由芝诺、毕达哥拉斯以至欧多克萨斯、欧几里得，不论是在哲学的思辨上或者在数学基础上都达到了很高的水平．但是，不但有这些纯思辨的考察，还有许多有待解决的实际问题，推动着科学的进步．从当时的生产力发展的水平以及与之相应的科学技术发展水平来看，静力学，流体的静力学，光学，特别是物体的机械运动，包括地上的和天上的物体的运动都是当时科学发展中的关键问题．其中一个重大问题是天体运行规律问题．

由古希腊的托勒密地心说以至后来哥白尼的日心说，以至布鲁诺被施以火刑，伽利略受到教廷的迫害，我们时常看到的是它的意识形态方面．但是这个问题的意义远不止此．现在每一个中学生都懂得了什么是相对运动，那么，日心说与地心说之差别无非是采用了不同的坐标系，何必大动干戈呢？但是能把问题“看穿”到这个地步，前提是要能用数学很好地刻画运动．要能追索星体在天空运动的轨迹，写出其运动方程式，而不是只满足于亚里士多德式的思辨，说什么物体下坠是因为向下落是物体“自然的本性”之类．这就不但不是依赖哲学，而恰好是能摆脱某种哲学，回到科学的道路上，回到观测、推算等等科学方法的道路上．不仅是天体运动的规律，还有关于运动学的基本概念如速度、加速度如何理解，什么是力，什么是离心力等等都是16—17世纪前后科学界关心的焦点．牛顿力学三大定律正是在这样的科学氛围中出现的．它一方面吸收了许多人，特别是伽利略、笛卡儿的成就，另一方面自然也是牛顿本人的努力和天才的结晶．所以牛顿说自己是站在巨人肩上是很有道理、很有见解的．我们特别要提出，牛顿深受欧几里得《几何原本》的影响．在他的巨著《自然哲学之数学原理》(1686，以下简称《原理》，中译本，武汉出版社，1992)一书中，他就把三大定律列为“运动的公理或定律”并由此推导出《原理》中所有的命题．这一点确实是牛顿的伟大贡献．

可是，这一切努力绝不是只具有理论的兴趣，而要实际得多．例如，由于资本主义向全世界的扩展，人们需要作长距离的航海，而确定船只在海洋中的位置自然是重要的．确定纬度比较容易，只要测定某一个恒星离天顶的角度即可，但确定经度却要困难得多．大约从16世纪起，人们就开始利用时差来确定经度．但为此，一是要有关于天体运动的准确知识．1675英王查理二世建立格林尼治天文台，就是为了这个目的．研究天体运动的人不仅有伽利略和牛顿，还有许多人，其中例如有哈雷(Edmund Halley，1656—1742，哈雷彗星的发现者)，他所作的星图是最精确的．二是要有一个好的钟．单摆是当时人们测定时间的基本工具．关于单摆的研究工作的基础是由惠更斯和胡克(Robert Hooke，1635—1703，曾任英国皇家学会秘书，与牛顿为了万有引力定律的“知识产

权”闹过不小的矛盾)奠定的.但是为了计算经度,摆的误差每天不得超过 2～3 秒,这在当时是很难达到的.由此还产生过一次灾难:1707 年,一支由 5 艘军舰组成的英国舰队,在海军元帅叔莱尔(Admiral Sir Clawdisley Shorell)率领下在直布罗陀大败法国舰队.可是在凯旋回国时遇大雾,因无法确定舰队准确位置而触礁,五艘军舰沉没了四艘,死亡水兵二千余人.在举国大哗之下,英国国会于 1714 年通过“经度法案”,悬赏 20 000 英镑(约合今天 $1 000 000),征求确定经度之法.最后由一位哈里逊(John Harrison, 1693—1776)在 1761 年造出了可以实用的海钟(当时称为 chronometer)而获奖.哈里逊没有读过大学,是一位极为能干的工匠.(以上见于 N. F. Lane, Problem Solving in Complex Society, AMS, Notices, Vol.44, May (1997), 580～584).总之,微积分发展的环境已与希腊时代和中世纪有了天壤之别.

关于天体的运动——当时还主要是太阳系中行星运动的规律.由于问题的重要性,人们已经进行了多年的观测,积累了大量观测数据.到 17 世纪初,开普勒以多年观测数据为基础,提出了著名的开普勒三定律.

Ⅰ.开普勒第一定律:行星运行的轨道是椭圆,太阳位于椭圆的一个焦点上.

Ⅱ.开普勒第二定律:联结太阳和某行星的线段称为行星的动径.若采用极坐标,令太阳在原点处,动径就是一个向量 $\boldsymbol{r}$.在相同时间内,$\boldsymbol{r}$ 扫过的扇形面积相同.简单的说法就说各行星的面积速度是一个常数(当然,不同的行星之面积速度是不同的常数).

Ⅲ.开普勒第三定律:各行星的周期各不相同,但是周期 $T$ 与轨道椭圆的长半轴 $a$ 之 3/2 次方成正比,即 $T/a^{3/2}=\text{const}$,注意,这个常数适用于各个行星.

开普勒三定律是唯象定律.它没有解释为什么恰好有这三个定律成立.牛顿的功绩在于指出了,它们其实是一个更深刻的定律:万有引力定律的推论.万有引力定律指出,若有二质点,质量各为 $m_1$ 与 $m_2$,设一个质点在极坐标原点处,另一质点的动径为 $\boldsymbol{r}$,则第一质点对另一质点必有引力,其方向与 $-\boldsymbol{r}$ 相同,而大小则与 $r^2$ 成反比,或者写成

$$\boldsymbol{F}=-G\frac{m_1m_2}{r^2}\frac{\boldsymbol{r}}{r},\quad r=|\boldsymbol{r}|.$$

$G$ 称为引力常数,$\boldsymbol{r}/r$ 表示 $\boldsymbol{r}$ 方向的单位向量.当然,按牛顿第三定律,另一质点也对第一质点有引力,大小与 $\boldsymbol{F}$ 相同,方向则相反,所以是 $-\boldsymbol{F}$.

牛顿是怎样把开普勒三定律与万有引力定律联系起来的呢?下面我们看一下牛顿在《原理》中怎样证明一个与此相关的命题:若一质点在有心力场中运动,则其面积速度必为常数(《原理》,第二章,命题 1,定理 1).但是这还只是由万有引力定理导出开普勒第二定律.该书第二章,命题 2,定理 2 还进一步证明了若面积速度为常数,则质点必是受有心力的作用.关于牛顿万有引力定律与开普勒三定律的关系我们将在后面详述,现在我们来看牛顿对上述命题的原证,其目的在于探讨牛顿的方法的要点是什么.

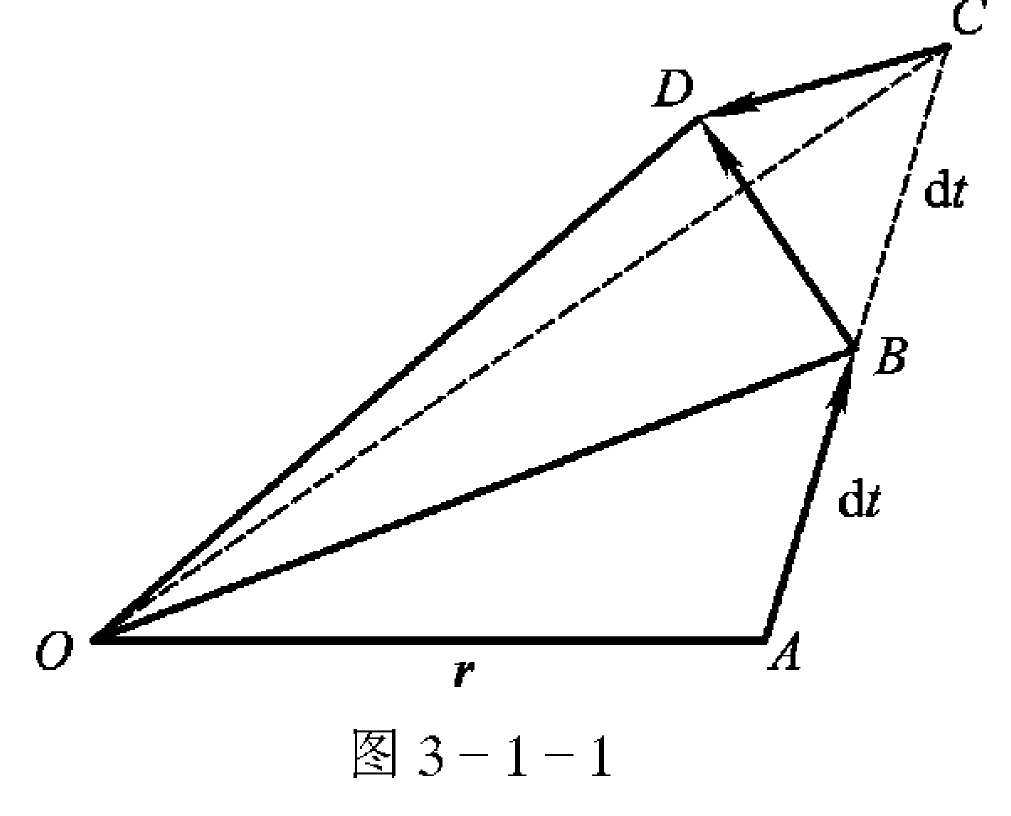


图 3-1-1

设有心力场的中心是原点 $O$.先设没有外力,这自然是有心力场的特例,即力之强度为 0.若质点初始位置在 $A$,经过一个无穷小时间 $\mathrm{d}t$ 后到了 $B$ 点.因为没有外力,按牛顿第一定律,该质点将沿直线 $AB$ 作等速运动.再过一段时间 $\mathrm{d}t$ 到达 $C$ 点.$ABC$ 是一条直线,而且 $AB=BC$.所以

$OB$ 是$\triangle OAC$之中线而$\triangle OAB=\triangle OBC$. 这就是说质点的动径$\overrightarrow{OA}$在 $dt$ 时间内扫过的面积相同(图 3-1-1).

若有向心力作用,则在 $B$ 点处有一个力沿$\overrightarrow{BO}$方向指向 $O$,因此在第二个时间段 $dt$ 内,质点还有一个与$\overrightarrow{BO}$方向平行的位移$\overrightarrow{CD}$. 这样质点从 $B$ 的位移将是$\overrightarrow{BC}$与$\overrightarrow{CD}$的合位移$\overrightarrow{BD}$. 而动径$\overrightarrow{OB}$在第二个时间段 $dt$ 内扫过的面积是$\triangle OBD$. 比较$\triangle OBC$与$\triangle OBD$,其底等长(均为 $OB$),而高又因顶点 $C$ 和 $D$ 位于 $OB$ 的平行直线 $CD$ 上,故又为等高,所以$\triangle OAB=\triangle OBC=\triangle OBD$. 因而定理证毕.

这个证明中最值得注意的是:牛顿的证明暗地里使用了一个假设,即在每一个时间段 $dt$ 里,有心力的方向与其开始瞬间的方向$\overrightarrow{BO}$一样,所以才得到面积速度可以用$\triangle OAB$与$\triangle OBD$表示的结论. 如果我们记质点动径(从$\overrightarrow{OA}$算起)扫过的面积为 $A(t)$(它当然是时间的函数),而在"无穷小"时段内 $dt$ 内,扫过的面积为 $dA(t)$,则在时刻 $t$

$$dA(t)=\triangle OAB+\text{高阶无穷小量},$$

在时刻 $t+dt$

$$dA(t+dt)=\triangle OBD+\text{高阶无穷小量},$$

注意到$\triangle OAB=\triangle OBD$,略去高阶无穷小量,并用 $dt$ 去除,即有

$$\frac{d^2A(t)}{dt^2}=0.$$

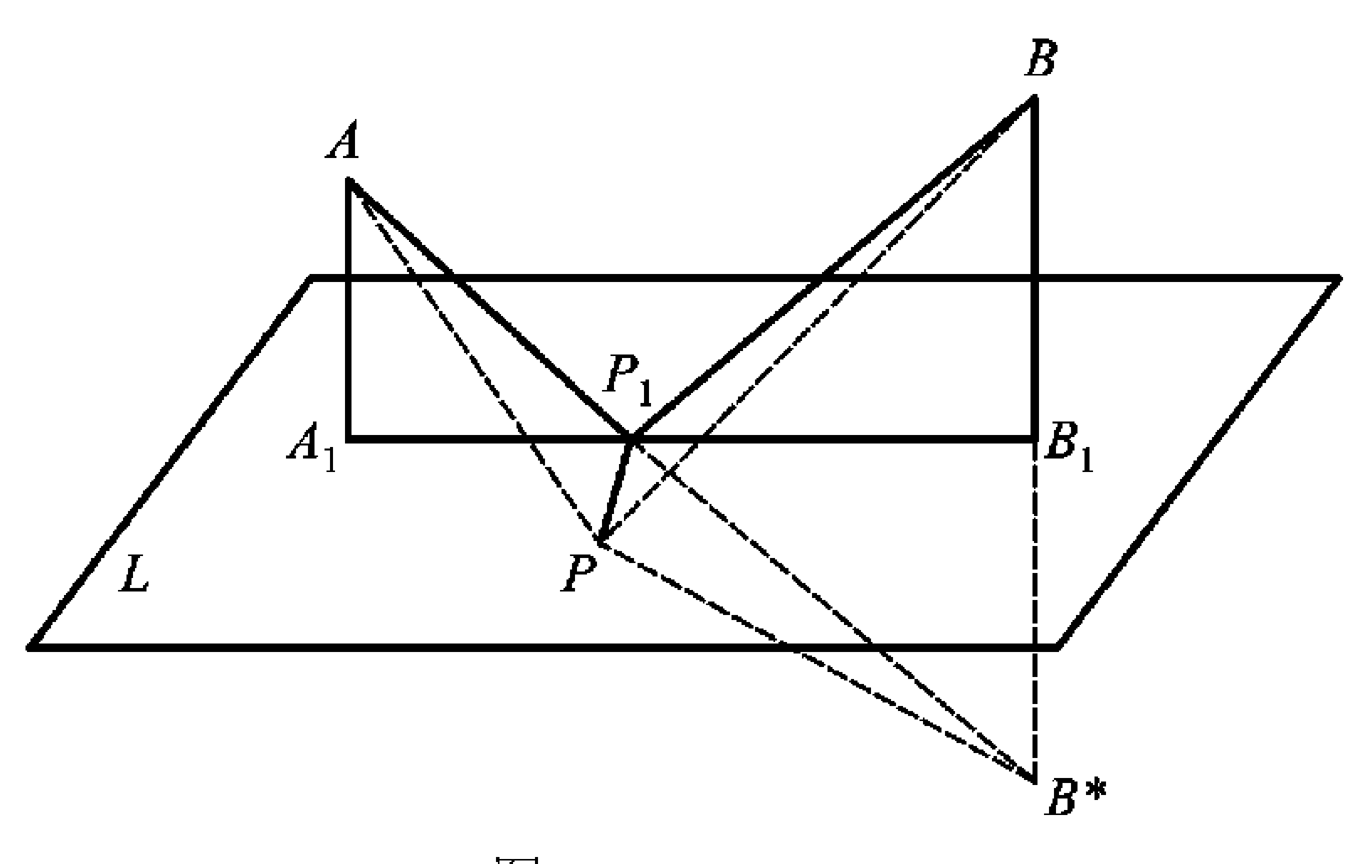



图 3-1-2

所以得知面积速度$\frac{dA(t)}{dt}$为常数.

下面再举一个光学中的例子,光学在 17 世纪也是一个重要的科学分支. 这当然是与望远镜和显微镜的发明分不开的. 而特别是望远镜,是天文研究最有力的工具. 于是几何光学兴起了(波动光学则晚到 19 世纪才流行起来),而 17—18 世纪的几乎所有大数学家、大物理学家(包括牛顿)对光学都作出了贡献. 我们特别要举出费马(Pierre de Fermat,1601—1665). 他其实也是微积分学的奠基者之一. 他的方法如下:例如,要考虑光的反射问题,设 $L$ 是一个平面镜面,光线从 $A$ 点射向 $L$ 并在点 $P$ 反射到 $B$ 点,问反射点 $P$ 的位置是什么? 当费马着手解决这个问题时,他已经有了一个著名的费马原理:光在一切可能的路径中必取耗时最短的路径. 这里要注意什么是"可能的路径"? 即联结 $A$ 与 $L$ 上一点 $P$ 再到 $B$ 的路径. 如果找到了这样一条路径,$AP$ 必为直线. 因为在镜面以外,光当然也走耗时最短因此长度也最短的路径——直线. 同理 $PB$ 也是直线. 我们可以看到,若由 $A$,$B$ 两点作 $L$ 之垂线,令垂足为 $A_1$,$B_1$,则 $AA_1$ 与 $BB_1$,决定一个与 $L$ 垂

直的平面$M$，$P$必在此平面上(图3－1－2).不然的话，由$P$向此平面$M$作垂线，令垂足为$P_1$，则$AP>AP_1$，同理$BP>BP_1$.从而$AP+PB>AP_1+P_1B$，而$APB$这条路径就不会是最短路径了.这样，问题就化成了平面问题：在平面$M$上有一直线$L$与线外同侧两点$A$与$B$，在$L$上求一点$P_1$使$AP_1+P_1B$达到最小值.

用对称点的方法求解这个问题是读者们熟知的.这个方法可以追溯到古希腊的海伦(Heron，大约生于公元1世纪)，但它不能推广到更复杂的问题，所以我们以折射问题为例说明费马的方法.设平面$M$上有一直线$L$，分$M$为上、下两半，在其中光速分别为$v_1$与$v_2$，在上、下半平面上各取一点$A$与$B$，问，若光线由$A$经$L$上某一点$P$折射后能到达$B$，$P$之位置何在(图3－1－3)?我们不妨设$P$之坐标为$x$，从而光走折线$APB$所需的时间是$x$的函数：

$$f(x)=\frac{AP}{v_1}+\frac{PB}{v_2}. \tag{1}$$

图3－1－3

费马的想法是，让$x$变动一个“无穷小量”$\mathrm{d}x$到$Q$，$Q$的坐标是$x+\mathrm{d}x$(费马当时用$E$或$e$来记$\mathrm{d}x$)，于是所需时间成为

$$f(x+\mathrm{d}x)=\frac{AQ}{v_1}+\frac{QB}{v_2}.$$

比较$f(x+\mathrm{d}x)$与$f(x)$.如果$f$在$x$点达到极小，则当向右移动($\mathrm{d}x>0$)时，$f(x+\mathrm{d}x)-f(x)\geqslant0$，就表示$f(x)$不会再有下降的可能；当$x$向左移动($\mathrm{d}x<0$)时，则又有$f(x+\mathrm{d}x)-f(x)\geqslant0$，就表示$f(x)$不会比$f(x+\mathrm{d}x)$更大了.同样，若$\mathrm{d}x>0$时，$f(x+\mathrm{d}x)-f(x)<0$，$\mathrm{d}x>0$时，$f(x+\mathrm{d}x)-f(x)<0$，也表明$f(x)$不是极小.因此，要使$f(x)$为极小，$f(x+\mathrm{d}x)-f(x)$不论$\mathrm{d}x$符号如何，总是非负的.我们就来计算$f(x+\mathrm{d}x)-f(x)$，具体说来就是要计算$(AQ-AP)+(BQ-BP)$.至此，一切都是显然的.费马的“诀窍”就在于计算$AQ-AP$时略去了“应该略去”的东西.过$P$点作$AQ$之垂线$PP_1$，如果$\mathrm{d}x$是无穷小，则由正弦定理，$\mathrm{d}\theta$也是同阶的“无穷小”.(注意，我们这里使用了现代的语言，在费马时代，人们还为无穷小量是不是不可分量等等而纠缠不清)所以$AP_1=AP\cos(\mathrm{d}\theta)=AP$(其中最多相差一个高阶无穷小量，用费马的记号，则是相差一个$E^2$差不多的东西.在费马看来，这个差不多的东西正是“应该略去的”东西.)因此

$$AQ-AP=P_1Q=PQ\cos(\theta-\mathrm{d}\theta)=\mathrm{d}x\cdot\cos\theta.$$

同理

$$QB-PB=-PQ_1=-PQ\cos\varphi=-\mathrm{d}x\cdot\cos\varphi.$$

这样

$$f(x+\mathrm{d}x)-f(x)=\left[\frac{\cos\theta}{v_1}-\frac{\cos\varphi}{v_2}\right]\mathrm{d}x.$$

这里最多相差一个高阶无穷小，用费马的记号是相差一个$E^2$的同类项.上式双方除以$\mathrm{d}x$即除以$E$，令$\mathrm{d}x\to0$(费马本人就干脆说令$E=0$)则费马推论出，折射点的位置$P$应适合$\cos\theta/v_1=\cos\varphi/v_2$.只有这样才能适合$x$点即$P$点为极小的条件.但在光学文献中我们通常用光线与$P$

点处的法线的夹角而不是与$L$的夹角来表示其方向.故$\theta=\frac{\pi}{2}-\theta_i$($\theta_i$称为入射角),$\varphi=\frac{\pi}{2}-\theta_r$,($\theta_r$称为折射角),就得到折射点的位置是

$$\frac{\sin\ \theta_i}{v_1}=\frac{\sin\ \theta_r}{v_2}.$$

在物理学中这叫做斯涅耳(Willibrord Snell)定律.其实笛卡儿也得到了这个定律.而且费马解决了希腊时代托勒密未能解决的问题.

用我们今天的语言来说,即令$\mathrm{d}x\to 0$而求极限,并令其为0:

$$\lim_{\mathrm{d}x\to 0}\frac{f(x+\mathrm{d}x)-f(x)}{\mathrm{d}x}=f'(x)=\left[\frac{\sin\ \theta_i}{v_1}-\frac{\sin\ \theta_r}{v_2}\right]=0.$$

但是在费马的时代人们还在为无穷小量是什么而困扰:费马说$E$是无穷小量,那么它究竟是不是0?如果是0,那么从一开始我们就停在$P$点,而不会前进到$Q$点,因此整个讨论就都无法进行了;如果不是0,为什么在经过许多计算后,在关键的一步上说"令$E=0$"?有了我们今天对无穷小量的理解,我们知道,无穷小量并不是"不可分量",而是趋于0的变量;并不是"令$E=0$",而是令$\mathrm{d}x\to 0$.高阶的无穷小量在除以$\mathrm{d}x$后,仍会趋向0,所以我们说是:丢掉高阶无穷小量.

以上我们通过实例说明了微分学的基本思想就是"丢掉高阶无穷小".要注意,当时的大数学家、大数学物理学家,不但是在某种社会经济发展和科学探索的需要的推动下工作,而且是在一定的文化、科学的传统下工作.从欧几里得以来,数学的严格性、合逻辑性的要求,并不如常人认为的那样是一种过分的苛求.牛顿写《原理》以《几何原本》为范本,就是这种表现.那么,什么是无穷小,它是不是零,诸如此类的问题自然成了他的最放不下心的事.所以,一方面牛顿自称"在数学中最微小的误差也不可忽略",另一方面,反对他的人就以此相讥:"高阶无穷小何以可以略去?"一直到19世纪中晚期,这场大争论才尘埃落定.微积分学才站在稳固的基础上,就是我们今天的教材的讲法,而略去高阶无穷小这个威力强大的方法也就得了无疑问的公认了.

**2.这个基本思想在数学物理中的展开**　自牛顿以后,这场争论如火如荼地进行了两百年.可是人们并不只是坐而论道,而是把牛顿的这个思想用于各个方面.因为牛顿既然对机械运动提出了一个包罗万象的理论,而且公认是人类思想史上对认识大自然的第一次伟大的综合(综合就是总结),人们当然就会把它用于机械运动的各个方面:流体、弹性体、声学……都纳入了科学家的视野,成了新生的数学武器——微积分学——的试验场.下面我们将要介绍"略去高阶无穷小"这个基本方法怎样帮助我们得到某些领域的基本规律的数学表述.但是这些规律的物理内容都比较简单.大体上都是例如质量守恒、动量守恒等等,因此都是牛顿力学的适用范围.

第一个例子是热传导方程.热现象本来在牛顿力学的视野之外.人们认真地研究热现象是19世纪的事.那时有了热力学,能量守恒定律内容的丰富性才开始展现.然后是统计物理,使统计的、随机的观点进入了物理学.因此,研究热现象应该从这个观点着眼.但是下面介绍的研究的出发点却是把热当作一种"流体"——热质说.从物理上说,它是很不够的,但是从数学上说却简单得多,而且带来了料想不及的丰硕成果.至于从概率的观点来描述热传导以至得出同样的方程,在第四章§4将简单介绍.

1811年巴黎科学院悬赏解决以下问题:"给出热的传播规律数学理论,并将理论结果与精密的实验结果相比较."应征得奖者是傅里叶(Jean-Baptiste Joseph Fourier,1768—1830).后来他把

自己关于这方面的结果写成一本书，即《热的解析理论》(Théorie Analytique de la Chaleur，中文译本，武汉出版社 1993 年出版). 这是一篇伟大的著作. 人们认为傅里叶的功绩有两个方面，首先他把物理问题用线性偏微分方程来表示，使得牛顿在《原理》一书中开创的道路远远扩展到了《原理》一书所规定的范围之外；其次，他为解决这些物理问题创造了一种数学理论，开创了“调和分析”理论，至今约二百年，长盛不衰，始终是数学主流的一部分. 这在数学史上是少见的. 本书将在第五章相当详尽地加以介绍.

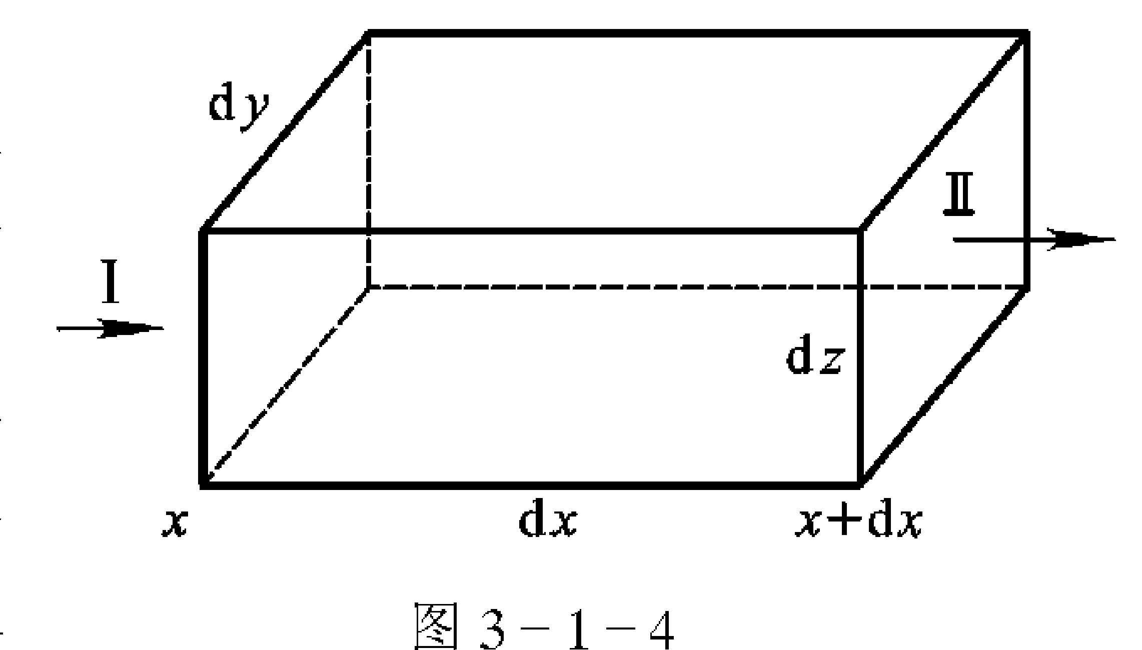



图 3-1-4

为简单起见，在热传导的区域中切出一块无穷小的长方体，其棱平行于坐标轴，而三边长度分别是无穷小 $dx, dy, dz$. 设热量通过其各个面进入这个无穷小区域(注意，关于热本质的了解要等到 19 世纪后半叶，而在傅里叶的时代，人们仍把热看成一种流体——热质，傅里叶的著作中也沿袭了这个观点)，如果我们记 $(x,y,z)$ 处时刻 $t$ 的温度为 $u(x,y,z,t)$，则例如图 3-1-4 断面Ⅰ处，依外法线方向(现在外法线方向即 $-x$ 方向)在时刻 $t$ 到 $t+dt$ 之间流入区域的“热量”应与温度的梯度、断面Ⅰ的面积以及时段长度成正比，因此应为

$$k\mathrm{d}y\mathrm{d}z\mathrm{d}t\,\frac{\partial u(x,y,z,t)}{\partial n}=-k\mathrm{d}y\mathrm{d}z\mathrm{d}t\,\frac{\partial u(x,y,z,t)}{\partial x}. \tag{2}$$

$k$ 是一个比例常数，称为传导系数. 一般说来，$k$ 是 $(x,y,z)$ 的函数，是表征介质的传热性质的，现在为简单起见设它是常数. 注意这里我们用了一个点 $(x,y,z)$、一个时刻 $t$ 的温度状况来表示一个断面、一段时间的温度状况，自然就会有误差，但由于断面面积 $dzdy$ 以及时段长度 $dt$ 均为无穷小，所以误差应为高阶无穷小，按我们上面的规定：高阶无穷小可以略去不计，这样才得到(2)式. 用同样的方法处理断面Ⅱ，不过注意，这时外法线方向即 $x$ 方向，从而$\frac{\partial u}{\partial n}=\frac{\partial u}{\partial x}$，由前所述知，经过断面Ⅱ在时间 $t$ 到 $t+dt$ 之间流入该无穷小区域的热量应是

$$k\mathrm{d}y\mathrm{d}z\mathrm{d}t\,\frac{\partial u(x+\mathrm{d}x,y,z,t)}{\partial x}. \tag{3}$$

(2)，(3)之和是

$$k\left[\frac{\partial u(x+\mathrm{d}x,y,z,t)}{\partial x}-\frac{\partial u(x,y,z,t)}{\partial x}\right]\mathrm{d}y\mathrm{d}z\mathrm{d}t=k\left[\frac{\partial^2 u(x,y,z,t)}{\partial x^2}\mathrm{d}x\right]\mathrm{d}y\mathrm{d}z\mathrm{d}t. \tag{4}$$

注意，在利用拉格朗日公式时，我们又略去了一个高阶无穷小量. 至此，$x$ 方向的热流已经清楚了. 用完全同样的方法处理 $y, z$ 方向的热流，我们将得到，在 $t$ 到 $t+dt$ 之间经过该无穷小区域之表面进入的热量总量是

$$\mathrm{d}Q=k\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t=k\Delta u(x,y,z,t)\mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t. \tag{5}$$

这里我们引入了一个记号 $\Delta$，我们称它为拉普拉斯算子：

$$\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2},$$

它可能是数学物理中“最重要的”算子了. 它的形状如此简单是否隐含了大自然的某种奥秘？是的，它表现了空间(爱因斯坦以前的空间)是均匀和各向同性的. 我们暂置不论，倒是 $dQ$ 的流入

将导致温度的改变，即由 $u(x,y,z,t)$ 变成 $u(x,y,z,t+\mathrm{d}t)$，温差是 $u(x,y,z,t+\mathrm{d}t)-u(x,y,z,t)=\frac{\partial u}{\partial t}\mathrm{d}t$. 这里我们又略去了高阶无穷小. 可是使单位质量的介质温度上升所需的热量应与质量以及温度增量成正比，比例常数 $c$ 称为比热. 现在该无穷小区域的介质之质量是 $\rho\mathrm{d}x\mathrm{d}y\mathrm{d}z$ （$\rho$ 是密度），所以又应有

$$\mathrm{d}Q=c\rho\frac{\partial u}{\partial t}\mathrm{d}t\mathrm{d}x\mathrm{d}y\mathrm{d}z. \tag{6}$$

比较(5)和(6)，立即得到温度 $u$ 所应适合的方程

$$\Delta u=\frac{c\rho}{k}\frac{\partial u}{\partial t}=\frac{1}{a^2}\frac{\partial u}{\partial t},\ a^2=k/c\rho. \tag{7}$$

这里我们又假设了 $c,\rho$ 都是常数.

方程(7)称为热传导方程. 可是值得注意的是，不仅是热传导现象由它来描述，而且扩散现象，渗流，化学反应乃至生物的繁殖与迁移等等也都是由它或由它衍生的方程来反映. 数学物理中的偏微分方程大多有此特点. 例如，波动方程不但能反映声波，而且也能刻画电磁波，以及一定条件下的水波、空气中的波等等. 定常现象则常用拉普拉斯方程

$$\Delta u=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}=0 \tag{8}$$

来刻画. 静电场的势，引力场的势等等都适合这样的方程.

现在应特别注意以下的事实. (7)是一个很一般的方程. 即是说，只要传热的物体 $\Omega$ 是均匀的，因此 $c$ 和 $h$ 都是常数，密度 $\rho$ 也是常数，则不论该物体是什么形状，处于什么条件下，总是这一个方程. 因此，在解决具体问题时，总要附加一些条件. 一是应该知道在初始时刻（设为 $t=0$）该物体内的温度分布，这样就有了一个附加条件

$$u\Big|_{t=0}=\varphi(x,y,z),\ (x,y,z)\in\Omega, \tag{9}$$

$\varphi(x,y,z)$ 是已知函数. (9)称为初始条件. 二是 $\Omega$ 边界上 $\partial\Omega$ 的情况也很重要. 这里可以有多种情况，例如知道在 $\partial\Omega$ 上的温度分布，这时有

$$u\Big|_{\partial\Omega}=\psi(x,y,z),\ (x,y,z)\in\partial\Omega, \tag{$10_1$}$$

另一种情况是假设 $\Omega$ 与外界没有热交换——绝热情况，则有

$$\frac{\partial u}{\partial n}\Big|_{\partial\Omega}=0, \tag{$10_2$}$$

$\boldsymbol{n}$ 是 $\partial\Omega$ 的法线方向. 还可能有其它情况. $(10_1)$ 与 $(10_2)$ 中只需取定一个就可以了，这称为边值条件. 至于一个方程应该附加什么样的条件，在有了这些附加条件下是否有解存在，是否只有一个解，这个解（或这些解）有什么性质，又如何求法，这些都是严重问题. 我们现在只想指出：一方面有一个方程（如(8)那样）来刻画一般的物理规律，另一方面又有一些附加条件，反映具体问题的具体情况，这是牛顿的伟大创造. 并不是说牛顿给出了初始条件、边值条件这些名词和概念，而是从他开始就这样研究问题了.

第二个例子是关于流体的动力学. 我们把流体看作连续介质（在一些极端情况下就不能这样看. 例如在大气层上层空气极端稀薄处，就需要把气体看成随机运动、碰撞的分子的集合. 这时就

不能采用连续介质的模型. 但是在牛顿的时代以及后来的 18—19 世纪当然不是这样). 这样一来, 流体力学就成了牛顿力学的延伸. 在这个模型下研究流体的基本方法是: 在流体中切出无穷小一块, 视作一个质点. 然后考虑质量守恒、动量守恒即可. 为什么不说能量守恒? 因为要讨论流体(特别是气体)中的能量守恒, 就至少要加上热力学的考虑. 这样一来, 关于内能, 熵……会出现一大堆新问题. 流体力学对于空气中的飞行, 如飞机、导弹等等, 当然是极其重要的. 到了 20 世纪中叶, 例如由于飞机速度越来越高, 这些热力学考虑是不可避免的. 在更高速度情况下, 还要考虑电磁场的影响, 化学反应的影响等等. 但是在 18—19 世纪, 还可以只限于机械能的考虑. 例如对船舶在水中航行, 通常都把水看成不可压缩流体, 这样只考虑机械能就够了. 不过即使在这时, 对流体的基本性质也要加上限制. 我们的限制是只考虑理想流体. 研究各种流体都少不了研究其速度场: $\boldsymbol{u}(x,y,z,t)=(u_1,u_2,u_3)$, 这是一个依赖于 $(x,y,z,t)$ 的向量场, 还有密度 $\rho(x,y,z,t)$ 以及压强 $p(x,y,z,t)$. (因为我们有意忽略热力学的考虑, 所以也略去了温度 $T(x,y,z,t)$ 的研究). 压强就是单位面积上所受的压力, 它本应是有方向的量(有方向的量不只是向量, 还有张量, 这一点详见第七章). 但我们现在取 $p=p(x,y,z,t)$ 为一标量, 其原因就在于我们只考虑理想流体. 理想流体就是没有黏性, 又不必考虑热传导的流体. 如果我们在流体内作一截面 $S$, 则在紧邻接着 $S$ 的 $P(x,y,z)$ 点, 其一侧 Ⅰ 的流体必向另一侧 Ⅱ 的流体施压力(图 3-1-5). 这个压力还依赖于 $S$ 的方向(不妨设为由 Ⅱ 伸向 Ⅰ 的法线 $\boldsymbol{n}$ 方向), 所以压力包含了两个方向要素, 在数学上我们用张量来描述它. 所谓理想流体就是不论取什么样的截面 $S$, 压力方向总在上述法线方向 $\boldsymbol{n}$ 上, 其大小也不随 $\boldsymbol{n}$ 改变. 这就是没有黏性作用的表现. 所以在理想流体中, 由 Ⅰ 向 Ⅱ 施加的压力应为

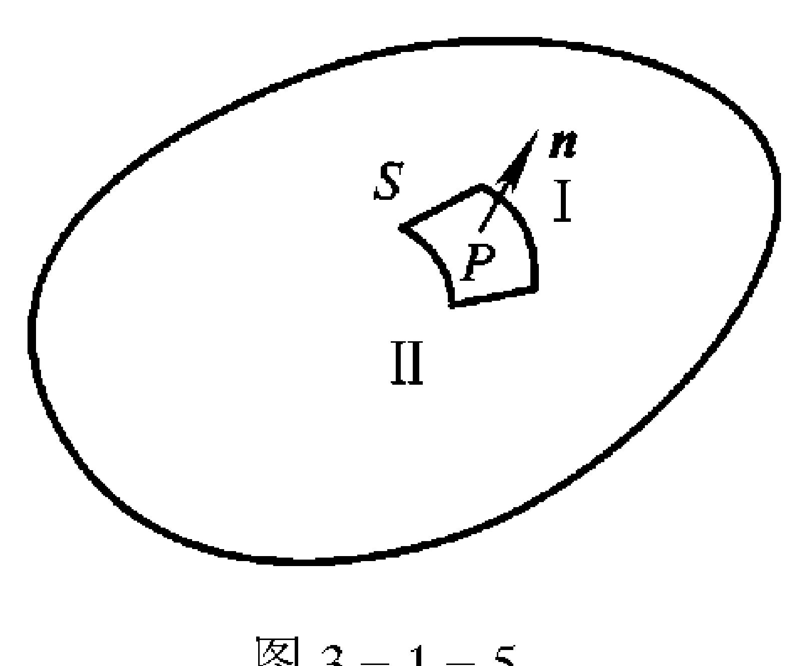


图 3-1-5

$$p(x,y,z,t)\cdot S\text{ 之面积}\cdot\boldsymbol{n}.$$

所以刻画理想流体各部分相互作用的内力, 只需要一个标量函数 $p=p(x,y,z,t)$ 就够了. 这就是规定压强为标量函数的原因.

现在在理想流体内切出其棱平行于坐标轴的无穷小块, 其三边之长分别是无穷小量 $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$. 我们先讨论质量守恒. 在 $(t,t+\mathrm{d}t)$ 时间段内从外面流入这个无穷小块的流体, 可以分别由经过 Ⅰ、Ⅱ……诸界面的流体计算得出(图 3-1-6). 例如经过 Ⅰ 的流体将成为一个边长为 $u_1\mathrm{d}t$ 而断面面积为 $\mathrm{d}y\mathrm{d}z$ 的柱子, 因此, 其质量为

$$(\rho u_1)(x,y,z,t)\mathrm{d}t\mathrm{d}y\mathrm{d}z.$$

同样, 经过 Ⅱ 流入的流体质量是 $-(\rho u_1)(x+\mathrm{d}x,y,z,t)\mathrm{d}t\mathrm{d}y\mathrm{d}z$, 所以由 $x$ 轴方向进入的流体总量是

$$-\frac{\partial}{\partial x}[\rho u_1(x,y,z,t)]\mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t.$$

而由全部边界进入的流体总量是

$$\mathrm{d}Q=-\left[\frac{\partial}{\partial x}(\rho u_1)+\frac{\partial}{\partial y}(\rho u_2)+\frac{\partial}{\partial z}(\rho u_3)\right](x,y,z,t)\mathrm{d}t\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

这些流体的进入将使该无穷小块中的密度由 $\rho(x,y,z,t)$ 变成 $\rho(x,y,z,t+\mathrm{d}t)$, 密度的增加所占用的流量就是 $\mathrm{d}Q$. 故在略去高阶无穷小后

$$\mathrm{d}Q=[\rho(x,y,z,t+\mathrm{d}t)-\rho(x,y,z,t)]\mathrm{d}x\mathrm{d}y\mathrm{d}z$$
$$=\frac{\partial\rho}{\partial t}(x,y,z,t)\mathrm{d}t\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

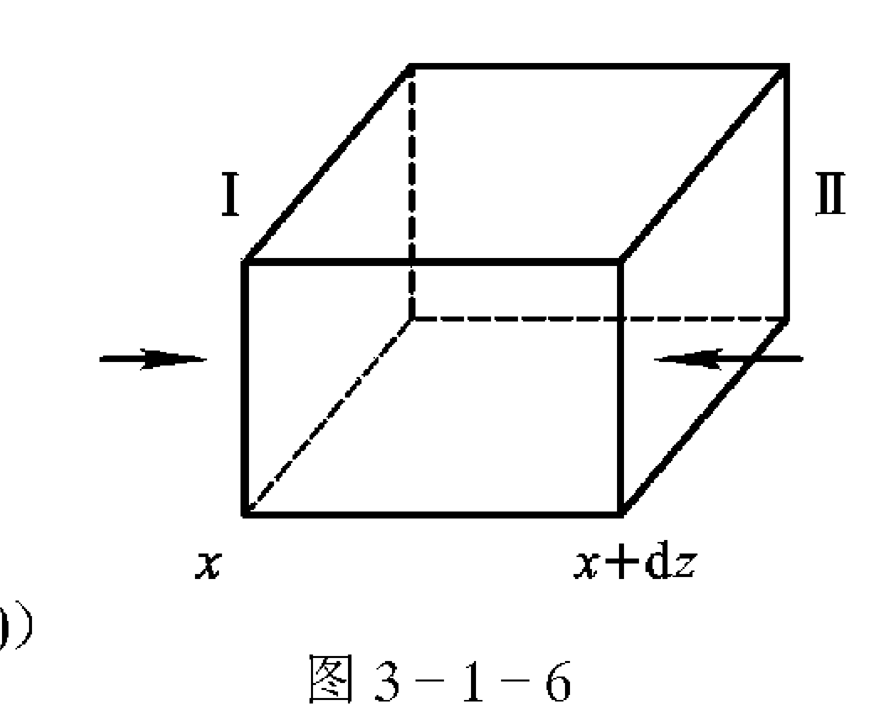


图 3-1-6

比较二式即得表征质量守恒的连续性方程：

$$\frac{\partial\rho}{\partial t}+\frac{\partial}{\partial x}(\rho u_1)+\frac{\partial}{\partial y}(\rho u_2)+\frac{\partial}{\partial z}(\rho u_3)$$
$$=\frac{\partial\rho}{\partial t}+\mathrm{div}(\rho\boldsymbol{u})=0. \tag{10}$$

再看动量守恒，我们也可以引入所谓动量密度流的概念，但这就涉及速度向量的张量积的概念，我们暂不假设读者都知道它，也不打算现在就花很大篇幅去讲解它(第七章中会讲到这里涉及的张量积)，因此就直截了当地把这个无穷小流体小块当成一个质点，其质量是 $\rho(x,y,z,t)\mathrm{d}x\mathrm{d}y\mathrm{d}z$，其速度为 $\boldsymbol{u}(x,y,z,t)$，然后直接应用牛顿第二定律. 其实这里已经包括了舍去高阶无穷小量的思想. 例如，一个无穷小元并不真正是一个点，因而其内部不一定是均匀的. 我们以 $(x,y,z,t)$ 处的密度作为此无穷小元内各点 $t$ 时的密度，本身就有一个误差，只不过这样算出的质量与真正的质量也就只差一个高阶无穷小. 下面计算加速度和外力时也都如此.

在计算加速度时要注意，速度 $\boldsymbol{u}(x,y,z,t)$ 中的 $x,y,z$ 都是 $t$ 的函数，$x=x(t)$，$y=y(t)$，$z=z(t)$ 即该质点的轨迹，所以在计算 $\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}$ 时，既要看到第四个变元 $t$ 在变化而 $x,y,z$ 不变时产生的加速度 $\frac{\partial\boldsymbol{u}}{\partial t}$(称为当地加速度)，又要看到因为 $(x,y,z)$ 随着时间变化而改变位置使 $\boldsymbol{u}$ 也得到一个加速度 $\frac{\partial\boldsymbol{u}}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial\boldsymbol{u}}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{\partial\boldsymbol{u}}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t}=u_1\frac{\partial\boldsymbol{u}}{\partial x}+u_2\frac{\partial\boldsymbol{u}}{\partial y}+u_3\frac{\partial\boldsymbol{u}}{\partial z}=\left(u_1\frac{\partial}{\partial x}+u_2\frac{\partial}{\partial y}+u_3\frac{\partial}{\partial z}\right)\boldsymbol{u}$ (称为对流加速度)，所以

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}=\text{当地加速度}+\text{对流加速度}$$
$$=\frac{\partial\boldsymbol{u}}{\partial t}+\left(u_1\frac{\partial}{\partial x}+u_2\frac{\partial}{\partial y}+u_3\frac{\partial}{\partial z}\right)\boldsymbol{u}=\frac{\partial\boldsymbol{u}}{\partial t}+(\boldsymbol{u}\cdot\mathbf{grad})\boldsymbol{u}. \tag{11}$$

最后看外力，外力有两个部分，一是流体之外来的力，例如重力. 又例如当流体是带电的，同时又处在电磁场中的电磁力，这时我们得到了电磁流体力学. 还有例如流体中有某种化学反应或燃烧、爆炸而产生的力. 我们记这种外力的质量密度(即作用在单位质量上的外力)为 $\boldsymbol{F}$. 第二部分是流体中这个无穷小元以外各部分对它的力，亦即压力，这一部分我们已计算了. 例如通过Ⅰ、Ⅱ两个面的作用力是

$$-p(x+\mathrm{d}x,y,z,t)\mathrm{d}y\mathrm{d}z+p(x,y,z,t)\mathrm{d}y\mathrm{d}z=-\frac{\partial p}{\partial x}\mathrm{d}x\mathrm{d}y\mathrm{d}z. \tag{12}$$

把这些结果综合起来，由牛顿第二定律：

$$\text{质量}\times\text{加速度}=\text{外力}$$

有

$$\rho\mathrm{d}x\mathrm{d}y\mathrm{d}z\,\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}=\rho\mathrm{d}x\mathrm{d}y\mathrm{d}z\cdot\boldsymbol{F}-\mathbf{grad}\,p\,\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

于是我们得到表现动量守恒的方程

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}=\frac{\partial \boldsymbol{u}}{\partial t}+(\boldsymbol{u}\cdot\mathbf{grad})\boldsymbol{u}=-\frac{1}{\rho}\mathbf{grad}\ p+\boldsymbol{F}. \tag{13}$$

(10)和(13)就是流体力学的基本方程组,如上所述,它们是不完备的,因为热力学的考虑都被忽略了.它是一个极广泛又极困难的领域的起点.

以上我们通过两个例子表明,应用微分学的基本思想于研究自然界的许多广泛领域,得到了描述其基本物理规律的微分方程.(这两个例子都是偏微分方程(PDE),下面我们将见到常微分方程(ODE)的例子.)微分方程是数学中范围极广泛、内容极丰富而且又极困难的领域.在许多自然科学分支中,它是基本的数学工具.我们可能以为略去高阶无穷小是产生这些微分方程的最重要的方法.实际上,这是由于我们现在研究的物理过程都比较简单,因而一些守恒律都变得很简单.随着科学的发展,我们需要解决的问题更深刻,所应用的数学工具也更深刻,这时,单只是"略去高阶无穷小"的方法就很不够了.我们将在本书最后介绍麦克斯韦方程组,以便了解数学的发展怎样帮助我们在更深的层次上认识自然界.

**3. 牛顿与开普勒** 牛顿写的《原理》是为了研究物体的运动,而所谓物体,上至日月星辰,下至苹果树上落下的苹果,无所不包.这样才看出《原理》是认识大自然的规律性的第一次伟大的综合.但是牛顿的三定律中讲的"物体",其实是指的质点.例如第一定律,牛顿的陈述原文如下:"每个物体都保持其静止、或匀速直线运动的状态,除非有外力作用于它迫使它改变那个状态".(《原理》中译本,13 页).然而牛顿紧接着举的例子却有陀螺,并指出陀螺各个部分的"凝聚力不断使之偏离直线运动,如果没有空气的阻碍,就不会停止旋转."可见,如果不是质点而是有内部构造,分成各部分的物体,如陀螺,牛顿也承认,其运动并不一定是匀速直线运动,而可以是旋转.把物体当成质点无非两个办法:一是如果空间极大,则哪怕是日月星辰,在浩渺无垠的太空中也只是沧海一粟的质点而已.二是物体有各部分,则各个部分若取得足够小,如上面两个例子中讲的无穷小的一块,小到对其内部结构,对其内部的不均匀性等等都可以忽略不计时,则其各部分也就看成一个质点,而物体整体则看成一个质点组.当然质点组中的质点总数可以是有限的,例如研究地球受太阳、月亮的作用而运动时,我们只需考虑由日、月、地三个质点形成的质点组,而其它行星和火星等等的影响都暂不考虑了,这就是著名的三体问题.这样,由牛顿的第二定律可以得出

$$m_i\frac{\mathrm{d}^2x_i}{\mathrm{d}t^2}=f_i\left(t,x,\frac{\mathrm{d}x}{\mathrm{d}t}\right),i=1,\cdots,N. \tag{14}$$

注意,这里的 $N$ 不一定是质点的个数.因为每一个质点可以有三个坐标,即相应于三个 $x_i(t)$.

牛顿遇到的问题通常并不是已知 $x_1(t),\cdots,x_N(t)$ 然后求它们的二阶导数,而通常是已知 $f_i\left(t,x,\frac{\mathrm{d}x}{\mathrm{d}t}\right)$,要求未知的 $x_i(t)$.这就是求解作为一个常微分方程(ODE)组的(14).正如前一段所说的,应该加上某些附加条件.从物理上看,只要知道这个质点组内各质点在初始时刻(例如可设为 $t=0$)的位置及速度就足以完全确定这个质点组在以后的运动状况.因此,我们所需的附加条件是

$$x_i(0)=x_i^0,\frac{\mathrm{d}x_i(0)}{\mathrm{d}t}=v_i^0. \tag{15}$$

(15)就是一个初始条件.在一个初始条件下求解(14),称为一个柯西问题或初值问题(initial value problem).以上讲的是有限质点组的情况.如果是无穷多个质点所成的质点组,如上面讲的连

续介质那样，就会得出偏微分方程组(PDE). ODE 和 PDE 是两个十分宽阔的数学领域. 而且二者由于提出的问题和处理方法不同，又形成很不相似的数学分支. 但不论如何，牛顿以后，在18—19世纪中，用数学方法研究大自然的数学规律，几乎就等同于研究微分方程.

牛顿既然已弄清楚了两个物体——质点之间的引力为

$$\boldsymbol{F}=-G\frac{m_1m_2}{r^2}\frac{\boldsymbol{r}}{r},\quad r=|\boldsymbol{r}|,\tag{16}$$

这里设第一个质点(其质量为 $m_1$)位于原点，$\boldsymbol{r}$ 则是由原点到第二个质点(质量 $m_2$)的向量(如果采用极坐标，$\boldsymbol{r}$ 就是第二个质点的动径)，$G$ 是引力常数，则为了求第二个质点的运动方程，只要将(16)代入(14)式并求解这个二阶常微分方程就行了. 这个问题称为开普勒问题，或称二体问题. 它是由瑞士数学家约翰·伯努利(Johann Beronulli, 1667—1748)①于1710年最早给出了完整的解答，这已是《原理》以后的23年了. 我们看见，现在已不需要再重复“略去高阶无穷小”这个思想，而只要应用在它的基础上建立起来的微积分的种种概念、方法和技巧就行了. 鉴于二体问题的解决很好地表明了牛顿的思想的力量，又是科学史上一件大事，我们用现在常见的微积分教材中的方法介绍一下它的解法.

我们记 $k=Gm_1$，并将第二个质点的质量 $m_2$ 就写成 $m$，于是(16)中的 $\boldsymbol{F}$ 成为 $\boldsymbol{F}=-\frac{km}{r^2}\frac{\boldsymbol{r}}{r}$，而第二个质点(以下就简称为质点，因为第一个质点已规定静止于原点处，或者说我们以第一个质点为参考系)的运动方程现在成为

$$\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}=-\frac{km}{r^2}\frac{\boldsymbol{r}}{r}.\tag{17}$$

$\boldsymbol{r}=(x,y,z)$ 是质点的坐标. 为简单起见，我们令 $m=1$ 上面的 ODE 组由三个 2 阶方程组成，相当于 6 个一阶 ODE. 这是一个相当大的方程组. 求解的基本思想是降阶，而降阶则通过寻找守恒律来实现. 如(17)这样的方程，力 $\boldsymbol{F}$ 的大小只依赖于 $r$，方向又是 $\boldsymbol{r}$，这种力称为有心力. 我们有

**定理 1** 质点在有心力作用下运动时，角动量守恒.

**证** 一个质点的角动量就是其动量 $m\dot{\boldsymbol{r}}$(本书遵从牛顿的记号，对时间求导用上加一点来表示)对原点的矩

$$\boldsymbol{M}(t)=m\boldsymbol{r}(t)\times\dot{\boldsymbol{r}}(t).$$

这里“×”表示向量积. 不论是标量函数的积，还是标量与向量之积，还是向量之间的标量积、向量积、混合积，求导时都服从莱布尼茨定律. 所以

$$\dot{\boldsymbol{M}}(t)=m\dot{\boldsymbol{r}}\times\dot{\boldsymbol{r}}+m\boldsymbol{r}\times\ddot{\boldsymbol{r}}=0.$$

第一项等于0是因为 $\dot{\boldsymbol{r}}$ 与 $\dot{\boldsymbol{r}}$ 平行，而两个平行向量的向量积一定为0. 第二项为0同样是因为 $\ddot{\boldsymbol{r}}$ 因是有心力而与 $\boldsymbol{r}$ 平行. 所以角动量 $\boldsymbol{M}$ 不随时间改变，定理证毕.

① 这里讲的是老约翰·伯努利，有时也称他是约翰·伯努利一世(Johann Ⅰ, Bernoulli). 他的儿子也叫约翰，所以常称为约翰·伯努利二世(Johann Ⅱ, Bernoulli, 1710—1790)，也是一位数学家，我们在变分学一节中将介绍他. 二世的哥哥丹尼尔·伯努利(Daniell Bernoulli, 1700—1782)也是一位大数学家. 对流体力学与气体分子运动论、偏微分方程等有大的贡献. 至于捷线问题，其实是一世和他的大哥雅各布一世(Jacob Ⅰ, Bernoulli, 1654—1705)开始研究的. 雅各布还是概率论的创始人之一. 伯努利一家的故事比杨家将真实多了又重要多了.

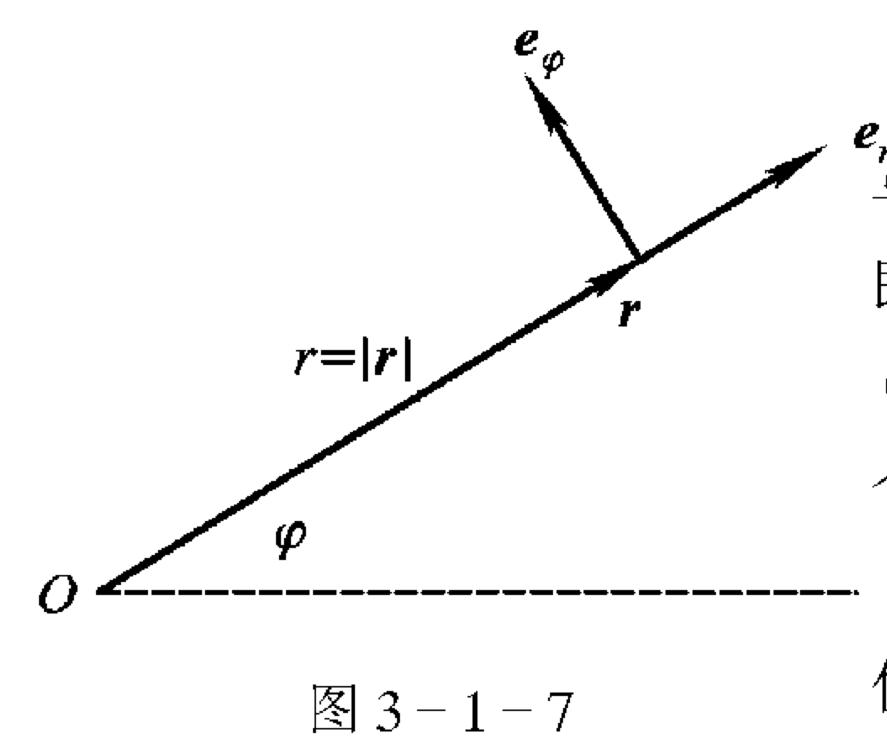


图 3-1-7

$\dot{M}(t)=0$ 不仅表示 $M(t)$ 大小不变而且方向也不变.这样,$r$ 与 $\dot{r}$ 恒在与 $M$ 正交的平面上.这样一来(17)就变成了平面问题,即 $r$ 是平面向量,而只有两个分量.不妨设它们就是 $r(t)=(x(t),y(t))$,于是(17)成了两个二阶 ODE 的方程组,相当于 4 个一阶 ODE.这就是降阶.

我们还可以应用 $M(t)$ 之大小也不变.既然开普勒问题已经化为平面问题,我们就可以引入极坐标系.于是在每一点 $(r,\varphi)$ 处可以引入两个单位向量 $e_r$——径向向量与 $e_\varphi$——横向向量,而且 $(e_r,e_\varphi)$ 取为右手系(图 3-1-7),而且

$$r=re_r+\varphi e_\varphi. \tag{18}$$

我们想在极坐标系下计算一下速度 $\dot{r}$.在第二章中讲三角函数时我们就讲过这个结果(但是使用的记号不同.现在的 $e_r$ 与 $e_\varphi$ 那时记作 $\rho$ 与 $\theta$),由于其重要性,我们再从角动量守恒角度再来算一次.

因为 $e_r$ 是单位向量,$e_r\cdot e_r=1$.对 $t$ 求导有

$$\dot{e}_r\cdot e_r+e_r\cdot\dot{e}_r=2e_r\cdot\dot{e}_r=0.$$

所以 $\dot{e}_r$ 与 $e_r$ 正交而与 $e_\varphi$ 平行.问题是同向还是反向?但是 $e_r$ 是绕单位圆旋转,所以 $\dot{e}_r=\dot{\varphi}e_\varphi$,同理 $\dot{e}_\varphi=-\dot{\varphi}e_r$:

$$\dot{e}_r=\dot{\varphi}e_\varphi,\dot{e}_\varphi=-\dot{\varphi}e_r. \tag{19}$$

这样,由于 $r=re_r$,我们立即有

$$\dot{r}=\dot{r}e_r+r\dot{e}_r=\dot{r}e_r+r\dot{\varphi}e_\varphi. \tag{20}$$

这就是说,速度向量有两个分量,一是径向的 $\dot{r}e_r$,另一是横向的 $r\dot{\varphi}e_\varphi$.利用它来计算角动量(以下均设质点的质量 $m=1$),有

$$\begin{aligned}M(t)=r\times\dot{r}&=\dot{r}(r\times e_r)+r^2\dot{\varphi}(e_r\times e_\varphi)\\&=r^2\dot{\varphi}(e_r\times e_\varphi).\end{aligned}$$

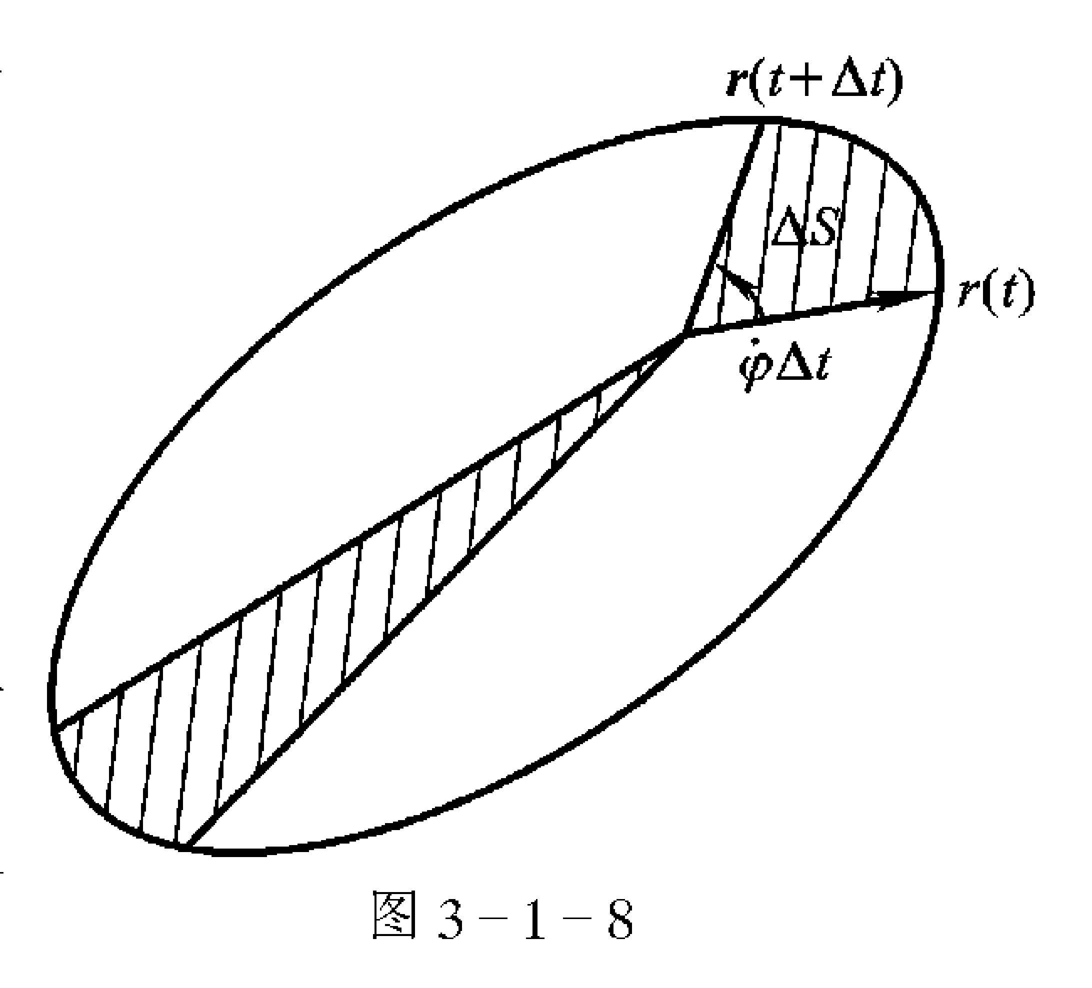


图 3-1-8

因为 $e_r\times e_\varphi$ 是 $M(t)$ 方向的单位向量(这由向量积的定义可知),所以 $M(t)$ 之"大小"是 $r^2\dot{\varphi}$.($|M(t)|\geqslant 0$ 而 $M(t)$ 之"大小"却因 $\dot{\varphi}$ 可能为负而可能小于 0),它是守恒的.

设动径 $r(t)$ 扫过的扇形(以原点为顶点)的面积为 $S(t)$.则由图 3-1-8 易见

$$\Delta S=S(t+\Delta t)-S(t)=\frac{1}{2}r^2\dot{\varphi}\Delta t+o(1)\Delta t.$$

所以"面积速度"

$$\frac{\mathrm{d}S}{\mathrm{d}t}=\frac{1}{2}r^2\dot{\varphi}=\frac{1}{2}M(t).$$

因为角动量不随时间变化，所以面积速度也不变. 这就是开普勒第二定律.

**定理 2（开普勒第二定律）**　有心力场中运动的质点恒有不随时间变化的面积速度.

这也就是前面讲的第一个例子.

为了看开普勒另外两个定律，我们再来看能量守恒. 首先注意，方程(17)中的力是有位能的. 其位能是

$$U(r)=-\frac{k}{r}.$$

这是因为牛顿的引力是 $-\mathbf{grad}\ U(r)$，事实上

$$-\frac{k}{r^2}\frac{\boldsymbol{r}}{r}=-\mathbf{grad}\ U(r)=k\,\mathbf{grad}\left(\frac{1}{r}\right).$$

但是我们要把(17)这个 2 维问题化为 1 维问题. 为此我们记住角动量“大小”的守恒值为 $M=r^2\dot{\varphi}$. 由(20)式对 $t$ 再求导，有

$$\begin{aligned}\ddot{\boldsymbol{r}}&=\ddot{r}\boldsymbol{e}_r+\dot{r}\dot{\boldsymbol{e}}_r+\dot{r}\dot{\varphi}\boldsymbol{e}_\varphi+r\ddot{\varphi}\boldsymbol{e}_\varphi+r\dot{\varphi}\dot{\boldsymbol{e}}_\varphi\\&=(\ddot{r}-r\dot{\varphi}^2)\boldsymbol{e}_r+(2\dot{r}\dot{\varphi}+r\ddot{\varphi})\boldsymbol{e}_\varphi=-\mathbf{grad}\ U(r).\end{aligned}$$

但是

$$\mathbf{grad}\ U(r)=\left(U'(r)\frac{x}{r},U'(r)\frac{y}{r}\right)=U'(r)\frac{\boldsymbol{r}}{r}=U'(r)\boldsymbol{e}_r,$$

所以有

$$\ddot{r}-r\dot{\varphi}^2=-U'(r),\quad 2\dot{r}\dot{\varphi}+r\ddot{\varphi}=0. \tag{21}$$

以角动量之守恒值代入，得到 $r$ 所适合的方程为

$$\ddot{r}=-\frac{\partial U}{\partial r}+r\frac{M^2}{r^4}=-\frac{\partial}{\partial r}\left(U+\frac{M^2}{2r^2}\right)=-\frac{\partial V}{\partial r}.$$

这里 $V(r)=U(r)+\dfrac{M^2}{2r^2}$称为有效位能.

(21)的后一式没有给我们新信息，因为若用 $r$ 乘此式两边，会得到

$$2r\dot{r}\dot{\varphi}+r^2\ddot{\varphi}=\frac{\mathrm{d}}{\mathrm{d}t}(r^2\dot{\varphi})=0,$$

亦即角动量之“大小”不变.

**定理 3**　服从(17)的质点的总能量是

$$\frac{1}{2}m\dot{\boldsymbol{r}}^2+mU=\frac{1}{2}m\dot{r}^2+mV.$$

**证**　由(20)

$$\dot{\boldsymbol{r}}^2=\dot{\boldsymbol{r}}\cdot\dot{\boldsymbol{r}}=\dot{r}^2+r^2\dot{\varphi}^2=\dot{r}^2+\frac{M^2}{r^2}.$$

所以

$$\frac{1}{2}\dot{\boldsymbol{r}}^2+U=\frac{1}{2}\dot{r}^2+\left(U+\frac{M^2}{2r^2}\right)=\frac{1}{2}\dot{r}^2+V.$$

能量守恒和角动量"大小"的守恒就可以用来进一步降阶并求出轨道.事实上,如果令单位质量上的总能量为 $E$(这是一个常数),则由上式有

$$\frac{\dot{r}^2}{2}+V=E \text{ 或 } \dot{r}=\sqrt{2(E-V)}.$$

又由角动量的大小为 $r^2\dot{\varphi}=M$,有

$$\dot{\varphi}=M/r^2.$$

所以

$$\frac{\mathrm{d}\varphi}{\mathrm{d}r}=\dot{\varphi}/\dot{r}=\frac{M/r^2}{\sqrt{2(E-V)}}. \tag{22}$$

这样就立刻可以得出

**定理 4**(**开普勒第一定律**)　若有心力之位能是 $U(r)=-\frac{k}{r}$,则质点运动轨道必为一椭圆,而原点是其一个焦点.

**证**　由(22)

$$\varphi=\int\frac{M/r^2\cdot\mathrm{d}r}{\sqrt{2(E-V)}}.$$

以 $V$ 的表达式 $V(r)=U(r)+\frac{M^2}{2r^2}=-\frac{k}{r}+\frac{M^2}{2r^2}$代入上式并作积分$\left(\text{需要作一个变量变换 } \rho=\frac{1}{r}\right)$,立即有

$$\varphi=\arccos\frac{\frac{M}{r}-\frac{k}{M}}{\sqrt{2E+k^2/M^2}}.$$

或者

$$r=p/(1+e\cos\varphi). \tag{23}$$

这就是椭圆的极坐标方程,或称焦方程.其中

$$p=M^2/k,\ e=\sqrt{1+2EM^2/k^2}.$$

$p$ 称为椭圆的焦半径,$e$ 称为其离心率.它们与长短半轴 $a,b$ 之关系是

$$e=\frac{c}{a}=\sqrt{a^2-b^2}/a,\quad p=a(1-e^2).$$

在由不定积分求 $\varphi$ 时还有一个积分常数.我们取它为 0 其实就是假设 $\varphi$ 要由长轴起算.$\varphi=0$ 时 $r$ 达到最小值 $p/(1+e)$,这就是近心点(近日点).定理证毕.

**注**　上面的积分给出的(23)式其实只是一般的圆锥曲线的焦方程.当 $e<1$ 时它是椭圆.$e>1$ 时是双曲线,而 $e=1$ 则是抛物线.但是 $e$ 的大小由 $E$ 决定.$E<0$ 时才可能 $e<1$,又因 $E=\frac{\dot{r}^2}{2}+V\geqslant V$,所以$\sqrt{2(E-V)}$总是有意义的实数.

此外还要注意,我们并没有完全解决开普勒问题,因为我们并没有得到 $\boldsymbol{r}(t)$,$\varphi(t)$的表达式,而只得到了 $r$ 与 $\varphi$ 之间的关系式(23).也就是说,我们只得知行星轨道为椭圆,但行星在某

一时刻之具体位置并不知道.但是这并不重要,因为例如由 $r^2\dot{\varphi}=M$,再积分就可以求出 $\varphi$ 为 $t$ 的函数,然后 $r$ 作为 $t$ 的函数也可以求出.

余下的只有开普勒第三定律了.由 $r$ 和 $\varphi$ 的表达式(23)即知当 $\varphi$ 运动过 $2\pi$ 以后,$r$ 就会进入下一个周期,如果时间由 $t=0$ 到 $t=T$ 能使 $\varphi$ 增加 $2\pi$,则 $T$ 自然是周期.在一个周期内动径扫过椭圆一次,而面积速度 $\frac{\mathrm{d}S}{\mathrm{d}t}$ 又是常数 $M/2$($M$ 是角动量的"大小"),所以

$$\int_0^T \frac{\mathrm{d}S}{\mathrm{d}t}\mathrm{d}t=\frac{1}{2}MT=\pi ab.$$

等式最右端是轨道椭圆的面积.现在我们就用上式来计算周期

$$T=2\pi ab/M.$$

注意到

$$\begin{aligned}a&=p/(1-e^2)=\frac{M^2}{k}\Big/\left[1-\left(1+\frac{2EM^2}{k^2}\right)\right]\\&=\frac{M^2}{k}\Big/\frac{2(-E)M^2}{k^2}=\frac{k}{2|E|}.\end{aligned}$$

这里出现 $|E|$ 是因为现在我们讨论的是椭圆轨道因此 $E<0$,从而 $-E=|E|$.此外,由离心率之定义 $e=c/a=\sqrt{a^2-b^2}/a$,从而 $b=a\sqrt{1-e^2}=\frac{k}{2|E|}\cdot\sqrt{\frac{2|E|M^2}{k^2}}=\frac{M}{\sqrt{2|E|}}$.所以

$$T=2\pi ab/M=2\pi k[2|E|]^{-3/2}=2\pi a^{3/2}k^{-\frac{1}{2}}. \tag{24}$$

**定理 5(开普勒第三定律)** 当上述质点之轨道为椭圆时,其周期与质点质量无关而只与其轨道长半轴之半立方 $a^{3/2}$ 成正比.

**证** 只需注意 $k=Gm_1=$(引力常数)×(第一质点质量)即得.

以上我们由牛顿的万有引力定律导出了开普勒三定律.但是实际上牛顿是由开普勒定律得到万有引力定律的.牛顿实际上的思考过程当然很复杂,其中涉及许多具体力学问题的解决,例如对离心力的理解等等,这里有许多过程已不可考.而在《原理》一书中,牛顿则用逻辑方法证明了万有引力定律确是开普勒三定律的必然结论.由于牛顿是用几何方法做这个工作,其中省略了"应该略去"的高阶无穷小量,就是说,他还是用的不成熟的微积分学,所以现在读起来相当困难.在微积分已经成熟了的今天,所需要做的大体上也就是把以上的推理"颠倒过来",这就容易多了.以下记号均与上面相同.

首先,既然承认质点的轨道是椭圆.所以这就是平面问题.不妨设质点(质量为 $m_2$,但 $m_2$ 下面记作 $m$)的运动发生在 $(x,y)$ 平面上,其轨道为 $(x(t),y(t))$.若使用极坐标,则该质点之轨道为 $(r(t),\varphi(t))$.另一质点则在原点处,其质量是 $m_1$.

先还是看角动量,由第二定律应有

$$\begin{aligned}0&=\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{r}(t)\times\dot{\boldsymbol{r}}(t))=\dot{\boldsymbol{r}}(t)\times\dot{\boldsymbol{r}}(t)+\frac{1}{m}\boldsymbol{r}(t)\times\boldsymbol{f}(t)\\&=\frac{1}{m}\boldsymbol{r}(t)\times\boldsymbol{f}(t).\end{aligned}$$

所以 $\boldsymbol{f}(t)\parallel\boldsymbol{r}(t)$.而 $\boldsymbol{f}=f(x,y)\boldsymbol{e}_r$ 是未知的外力.下面我们来求 $f(x,y)$,并简记为 $f$.

由 $\boldsymbol{r}\times\dot{\boldsymbol{r}}$之大小也不变,即有

$$\boldsymbol{r}(t)\times[\dot{r}(t)\boldsymbol{e}_r+r\dot{\varphi}\boldsymbol{e}_\varphi]=r^2\dot{\varphi}(\boldsymbol{e}_r\times\boldsymbol{e}_\varphi)$$

之大小不变.所以

$$r^2\dot{\varphi}=M\neq 0. \tag{25}$$

(若 $M=0$,则 $\dot{\varphi}\equiv 0$ 而质点不可能绕椭圆运动).由此可知 $\dot{\varphi}\neq 0$.

为了求 $f(x,y)$,我们需要考虑质点的动能

$$\begin{aligned}K&=\frac{m}{2}\dot{\boldsymbol{r}}\cdot\dot{\boldsymbol{r}}=\frac{m}{2}(\dot{r}^2+r^2\dot{\varphi}^2)=\frac{m}{2}\left[\dot{r}^2+\frac{(r^2\dot{\varphi})^2}{r^2}\right]\\&=\frac{m}{2}\left[\dot{r}^2+\frac{M^2}{r^2}\right].\end{aligned}$$

双方对 $t$ 求导,有

$$m\dot{\boldsymbol{r}}\ddot{\boldsymbol{r}}=m\dot{r}\left(\ddot{r}-\frac{M^2}{r^3}\right). \qquad (\text{这里 }\dot{\boldsymbol{r}}\text{与 }\dot{r}\text{ 不同.})$$

以 $m\ddot{\boldsymbol{r}}=f\boldsymbol{e}_r$,$\dot{\boldsymbol{r}}=\dot{r}\boldsymbol{e}_r+r\dot{\varphi}\boldsymbol{e}_\varphi$,代入即有

$$f\dot{r}=m\dot{r}\left(\ddot{r}-\frac{M^2}{r^3}\right).$$

我们在下面将证明 $\dot{r}$可以约去,于是有

$$f=m\left(\ddot{r}-\frac{M^2}{r^3}\right). \tag{26}$$

为了计算 $\ddot{r}$,我们要利用椭圆轨道的焦方程

$$r(1+e\cos\varphi)=r+ex=p.$$

双方对 $t$ 求导二次.于是有

$$0=\ddot{r}+e\ddot{x}=\ddot{r}+\frac{e}{m}(\boldsymbol{f}\text{ 之 }x\text{ 分量})=\ddot{r}+\frac{e}{m}f(x,y)\cos\varphi.$$

即

$$\ddot{r}=-\frac{e}{m}f(x,y)\cos\varphi.$$

代入(26),并且注意到椭圆的焦方程 $r(1+e\cos\varphi)=p$,即有

$$f(x,y)=-\frac{m}{p}M^2\cdot\frac{1}{r^2},$$

而外力

$$\boldsymbol{f}=-\frac{mM^2}{p}\cdot\frac{1}{r^2}\boldsymbol{e}_r. \tag{27}$$

这就是说该质点受到的引力是有心的,而且与 $r^2$ 成反比.为什么上面提到可以约去 $\dot{r}$? 事实上,把椭圆方程对 $t$ 求导,将有

$$\dot{r}(1+e\cos\varphi)-er\sin\varphi\cdot\dot{\varphi}=0.$$

但上面已说了 $\dot{\varphi}=\frac{M}{r^2}\neq 0$，即 $\varphi=0,\pi$ 以外恒有 $\dot{r}\neq 0$. 因此除了在 $\varphi=0,\pi$（即近心点与远心点）外 (27) 恒成立. 求极限即知 (27) 在整个轨道上都成立.

现在要问，怎样求引力常数. 这就需要开普勒第三定律，令质点的周期为 $T$，和前面的证明一样

$$MT=\int_0^T r^2\dot{\varphi}\mathrm{d}t=\int_0^{2\pi} r^2\mathrm{d}\varphi=2\pi ab.$$

所以 (27) 中的常数 $M^2/p$ 是

$$\frac{M^2}{p}=\frac{4\pi^2a^2b^2}{T^2}\cdot\frac{a}{b^2}=\frac{4\pi a^3}{T^2}.$$

这里我们利用了 $p=a(1-e^2)=\frac{b^2}{a}$. 由开普勒第三定律，上式右方是一个适用于一切质点的常数. 我们再恢复最早的记号，即位于原点的质点质量为 $m_1$，而运动质点的质量为 $m_2$（即 $m=m_2$），把常数 $M^2/p$ 写成 $Gm_1$，$G$ 之大小与 $m_1$，$m_2$ 无关，称为引力常数，而 (27) 成为

$$\boldsymbol{f}=-\frac{Gm_1m_2}{r^2}\boldsymbol{e}_r. \tag{28}$$

这就是牛顿的万有引力定律.

## §2 什么是微分？

**1. 线性化** 现在进一步分析上一节的两个例子.

在光的折射问题中，我们发现光走过折线 $APB$（图 3-1-3）所需的时间 $f(x)$（上节 (1) 式）适合以下关系式

$$f(x+\mathrm{d}x)-f(x)=\left[\frac{\cos\theta}{v_1}-\frac{\cos\varphi}{v_2}\right]\mathrm{d}x+\text{高阶无穷小量}.$$

对于行星绕日问题，若记动径扫过的扇形面积为 $S(t)$（假设 $t=0$ 时动径位置在 $OA$，参看图 3-1-1），则在有引力存在情况下，在 $\mathrm{d}t$ 时间内，扫过的面积应是 $S(t+\mathrm{d}t)-S(t)$，而且除了一个高阶无穷小以外，这个差应为扇形 $OBD$（$\overrightarrow{OB}$ 为动径在时刻 $t$ 的位置），其顶角为 $\mathrm{d}\varphi=\angle BOD$，半径为 $r$，因此面积为 $\frac{1}{2}r^2\mathrm{d}\varphi=\frac{1}{2}r^2\dot{\varphi}\mathrm{d}t$，于是我们又有

$$S(t+\mathrm{d}t)-S(t)=\frac{1}{2}r^2\dot{\varphi}\mathrm{d}t+\text{高阶无穷小量}.$$

在这两个例子中，我们都是把注意力放在右方第一项上，在折射问题中，由费马原理得出

$$\frac{\cos\theta}{v_1}-\frac{\cos\varphi}{v_2}=0,$$

由此即得关于折射的斯涅尔定律，在行星运动问题中，我们得到了瞬时的面积速度 $\frac{\mathrm{d}S(t)}{\mathrm{d}t}=\frac{1}{2}r^2\dot{\varphi}=$ 常数.

概括这两个例子所用的数学方法，则都是对可求导的函数 $f(x)$ 使用了以下的公式

$$f(x+h)-f(x)=f'(x)h+o(h). \tag{1}$$

在所有关于微积分的教本中都称右方第一项为 $f(x)$ 在 $x$ 点的微分：

$$\mathrm{d}f(x)=f'(x)h=f'(x)\mathrm{d}x. \tag{2}$$

然后就是一连串无休止的争论：$\mathrm{d}f$ 是不是无穷小量，$\mathrm{d}x$ 是不是无穷小量？无论如何，把 $h$ 写成 $\mathrm{d}x$ 总似乎有某种暗示：对于相信 $h$ 是无穷小的人，把它写成 $\mathrm{d}x$ 似乎是暗示了它确实是无穷小量，否则从哪里谈得起“高阶无穷小”？对于为 $\mathrm{d}x$ 是无穷小所带来的神秘气氛而迷惑的人来说，把 $\mathrm{d}x$ 写成 $h$ 又似乎是在暗示 $\mathrm{d}x$ 其实也就是普通的量，而可以如处理通常的数一样去对待它. 特别是要注意 $h\neq 0$（否则例如在行星问题中，动径被永远地冻结在时刻 $t$ 的位置不动了，因此也就无所谓“扫过一个扇形”了），所以可以用 $h$ 作分母去作除法了. 而真正给出“致命一击”的仍然是伯克莱大主教：你们不是一再以数学的严格性而自傲吗？牛顿自己就说“在数学中最微小的误差也不可以忽略”，那么为什么又把高阶无穷小量丢掉了呢？费马先是把 $h$ 写成 $E$，还说明了 $E\neq 0$，可是 $E^2$ 却被他丢掉了. 这些问题困惑着三百年前的牛顿直到今天的大学生.

然而由牛顿的时代到 19 世纪晚期数学的进步都表明了，上面说的问题根本不是问题. 读者可能以为这是言过其实，我们不妨读一下魏尔斯特拉斯 1861 年在德国柏林皇家工科大学讲课的笔记. 下面是他的学生，著名数学家施瓦茨（H. A. Schwarz，我们常用的施瓦茨不等式就出自他手）的记录. 引文录自李文林编《数学珍宝》，科学出版社，1998 年，684～685 页，凡下加波线的都是魏尔斯特拉斯的原话，其余是本书作者的说明）：

“当 $x$ 变至 $x+h$ 时函数 $f(x)$ 所产生的全改变量 $f(x+h)-f(x)$ 一般（当 $f$ 可求导时就行——本书作者注）可分解为两个部分，第一部分正比于自变量的改变量 $h$，因此它由 $h$ 和不依赖于 $h$ 的一个因子（即关于 $h$ 为常量）（按：这里指(1)中的 $f'(x)$——本书作者注）组成，从而它当 $h$ 变为无穷小时也变为无穷小（魏尔斯特拉斯这里并没有把无穷小当成一个“怪物”，更不讨论它是否“不可分量”等等. 在这个笔记前一点，他就对“变为无穷小”下了定义：“如果一个量的绝对值能变得小于任意选定的无论怎样小的量，则我们说它能变为无穷小”，例如 $|x|<\varepsilon$. $\varepsilon$—$\delta$ 的记号就是他的创造. 关于 $\varepsilon$—$\delta$ 第一次清楚的表述也见于这份讲课的记录. 见同书 683 页——本书作者注），或随 $h$ 同时变为无穷小. 然而，另一部分当 $h$ 变为无穷小时不仅自身变为无穷小（即当除以 $h$ 后仍变为无穷小——本书作者注）.

……

函数全改变量的正比于自变量的改变量的第一部分，称为微分改变量或微分，并以刻画其特征的 $d$ 置于函数前表示，而前缀 $\Delta$ 则表示全改变量（这说明 $\mathrm{d}f$ 与 $\Delta f$ 只是两个符号，使用了这个记号后，(1)将表示为

$$\Delta f=\mathrm{d}f+o(h). \tag{3}$$

这里没有任何一点神秘色彩，也没有丢掉任何一个不为 0 的量——本书作者注）. 对 $h$ 我们也类似地写为 $\mathrm{d}x$，因为 $x$ 的最简单的函数是 $x$ 自身，而 $\mathrm{d}x$ 是完全不依赖于 $x$ 的能变为无穷小的量.（就是说，若令 $f(x)\equiv x$，这是 $x$ 的最简单的函数. 对于这个函数 $o(h)\equiv 0$. 其实，要想 $o(h)\equiv 0$，必要充分条件是 $f(x)\equiv Ax+C$，$A$ 和 $C$ 是任意常数. 对于这个函数 $\Delta f=f(x+h)-f(x)=A(x+h)-Ax=Ah$. 另一方面，因 $o(h)\equiv 0$，故由(3)又有 $\Delta x=\mathrm{d}x$. 总之，对于自变量 $x$ 而言，$\Delta x=\mathrm{d}x=h$，这样就有了(2)式. 而且，不存在 $\mathrm{d}x$ 是不是无穷小量的问题，它只

是能变为无穷小的量，即可以趋于零的量．十分重要的是，$dx$ 完全不依赖于 $x$，这一点我们在正文中要详细地讲解——本书作者注）．$dx$ 或 $h$ 取得越小，微分改变量与全改变量的差异越少，通过减小 $dx$ 能使此差别小于每个无论怎样小的量，因此人们定义微分为函数当其自变量改变一无穷小量时所产生的次变量．”

这一段话把古典的，即通常微积分教材中所讲的微分究竟是什么说得再清楚不过了．概括起来就是：

1．$\Delta f$ 可分为两部分，一部分与 $dx=h$ 成线性关系的称微分，记作 $df$；另一部分对 $h$ 是高阶无穷小量，这里什么都没有舍去．所以 $df$ 是 $\Delta f$ 的线性部分．许多书上都爱说“主要部分”，这至少是不准确的．因为例如表达式 $3h+h^2$ 比之 $h$，自然 $3h$ 是主要部分，因为 $3h/(3h+h^2)\to 1\ (h\to 0)$，而另一部分 $h^2$ 则有 $h^2/(3h+h^2)\to 0\ (h\to 0)$．但是它也可写为 $\left(3h+\frac{h^2}{2}\right)+\frac{h^2}{2}$，仍有 $\left(3h+\frac{h^2}{2}\right)\Big/(3h+h^2)\to 1$，$\frac{h^2}{2}\Big/(3h+h^2)\to 0$．那么为什么不说 $3h+\frac{h^2}{2}$ 是主要部分呢？再说，若 $f'$ 在 $x$ 点为 0，则由(1)式在此点 $f(x+h)-f(x)=o(h)$，而我们不能说 0 是 $\Delta f$ 的主要部分．线性部分这种说法是大有深意的．

2．$\Delta f$，$df$，$dx$ 无非是记号，它们都是普通的量．使用这些为了刻画它们是怎样来的，这里面没有任何神秘的东西．

3．有些时候我们要令 $h$ 很小，这时可以看到 $df$ 也很小，误差项 $o(h)$ 对某个 $h$，哪怕是很小的 $h$，都不一定在数值上更小，但一般说来会是更小的．在有必要比较 $f(x+h)-f(x)$ 与 $f'(x)h$ 时，确实需要让 $h$ 取很小的值，再说 $f(x)$ 可能只定义在某个区间 $[a,b]$ 上，若 $h$ 不是很小，$x+h$ 就会越出这个区间．特别是，若 $x=b$，必需令 $h\leqslant 0$，否则 $f(x+h)=f(b+h)$ 根本没有定义．此外，在需要比较 $f(x+h)-f(x)$ 与 $f'(x)h$ 作为无穷小量的阶时才有需要使 $h=dx$ 为无穷小量．正因为在实际使用起微分概念时，时常要让 $h$ 很小或甚至 $h\to 0$，这样一种习惯的力量使得“人们定义微分为函数当其自变量改变一无穷小量时所产生的改变量”．而作为一个数学概念 $df$ 就是 $\Delta f$ 的线性部分，即“正比于自变量的改变量 $h$”的那一部分：$f'(x)h$．这里对 $h$ 之大小没有任何限制．

然而一切数学概念凡十分重要的，如微分等等，都一定是随历史发展的．微分的现代的定义与上述古典定义之差别在于，现在把微分看作一个映射．于是若有一个定义在区间 $(a,b)$ 上的函数 $f(x)$，或者说有一个映射 $f:(a,b)\to\mathbf{R}$，如果 $f(x)$ 在 $(a,b)$ 上可求导，我们就给出

**定义 1** 对上述 $f(x)$ 以及 $x\in(a,b)$ 若有一线性映射 $A:\mathbf{R}\to\mathbf{R},h\mapsto Ah$，使得

$$f(x+h)-f(x)=Ah+o(h), \tag{4}$$

则映射 $A$ 称为 $f$ 在 $x$ 点的微分．

由函数导数之定义即知，若令 $A=f'(x)$，则这个 $A$ 就是微分．这里注意几点：

1．现在微分定义为一个算子、一个运算或称为一个映射．现在，$f'(x)$ 并不表示一个数，而是表示乘以 $f'(x)$ 这个数，因此是一个映射．这个映射称为原映射 $f$ 的切映射．古典定义把切映射作用于 $h\in\mathbf{R}$ 之结果称为微分，现在则把切映射本身称为微分．因此古典的微分 $df=Ah=f'(x)h$，现在的微分则是 $df=f'(x)$．(我们时常还是保留 $df$ 这个记号，这又会造成相当多的麻烦)．二者的区别有如函数值与函数关系．

2．把 $f$ 也看成映射，于是(4)就表示，用切映射 $f'(x)$ 去代替原来的映射 $f$．这个过程就称

为线性化.这是数学物理一个根本的处理问题的方法.

3. 在以上的讨论中我们必须把 $x$ 与 $h$ 分开.魏尔斯特拉斯的那一段话中就提到 $\mathrm{d}x=h$ 是完全不依赖于 $x$ 的.现在从几何上来看这个问题.$y=f(x)$ 是一条曲线,它定义在 $\mathbf{R}^1$——$x$ 轴——上的开区间 $(a,b)$ 上,但是它在 $P$ 点(自变量为 $x$ 处)有一切线.这切线可以无限延伸.其定义域是 $-\infty<h<+\infty$,即图 3-2-1 上的斜线,不过我们把它的原点 $h=0$ 移到了 $x$ 处.曲线与切线都是映射.前者是定义在 $(a,b)$ 上的非线性映射.后者是定义在 $-\infty<h<+\infty$ 上的线性映射.我们在图上画出了它们的图像(graph).这两个映射之差当 $h\to 0$ 时是 $h$ 的高阶无穷小.曲线的定义域 $(a,b)$ 是 $\mathbf{R}^1$——$x$ 轴——的一个开集,切线的定义域则是整个 $\mathbf{R}^1$ 空间 $-\infty<h<+\infty$,我们称此空间为曲线在 $P$ 点的切空间,记作 $T_P$.我们对上面用的图像二字还要作一些解释,若有一个将 $E$ 的子集 $A$(称为定义域)映入 $F$ 内的映射 $\Phi: A\subset E\to F$,所谓图像就是乘积集合 $E\times F$ 的子集 $\mathrm{graph}(\Phi)=\{(x,y); x\in A, y=\Phi(x)\}$.所以上图中有两个映射,一是 $f:(a,b)\to\{y$ 轴$\}$,其自变量用 $x$ 表示;一是切线 $SPT$:$\{-\infty<h<+\infty\}\to\{y$ 轴$\}$,其自变量用 $h$ 表示.它的图像一是曲线 $RQ$,一是直线 $PT$.我们还说现在有两个空间,一是曲线 $\widehat{RPQ}$(或者说它的定义域 $(a,b)$),称为底空间,二是直线 $SPT$(或 $h$ 轴 $\mathbf{R}^1$)称为切空间.有两个映射 $f:(a,b)\to\mathbf{R}^1$ 和 $\mathrm{d}f: R_h\to\mathbf{R}^1$,后者称为前者在 $x$ 点的切映射.魏尔斯特拉斯说的"$h$ 完全不依赖于 $x$"就是说我们研究切映射 $\mathrm{d}f(x)$ 时,要以 $h$ 为自变量,而 $x$ 则视为固定的,最多也是看作一个参数.我们常把 $x$ 与 $h$ 的变化混在一起分不清是因为现在我们讨论的仅是最简单的情况.要研究相应于 $h$ 的函数的全改变量 $f(x+h)-f(x)$ 与 $h$ 之关系,这样就必须而且也能够把这两个不同的变量加起来,甚至给人一个印象即切空间是一个仿射空间:其原点可以任意放到一个位置上(现在是放到底空间的 $x$ 点),而在研究一般的映射与切映射之间的关系时,就不一定(也不需要)能把两个不同的变量加起来,我们研究的也不是 $f(x+h)-f(x)$ 这样的问题.

图 3-2-1

以上我们是固定了 $x$ 来看切映射 $\mathrm{d}f(x)$ 的.如果 $x$ 变动了又如何?例如令 $x$ 变到另一点 $x'$,$P$ 点变成了 $P'$,则切空间(图 3-2-2 上的虚线)也要变,其原点对着 $x'$ 处,我们把它看成与 $P$ 点的切空间不同的一维空间.这样一来,联系着一小段曲线 $RPQ$,就有许许多多的切空间,各视为互不相干的,这一族切空间称为 $\widehat{RPQ}$ 的切丛.不过我们要注意,一个线性空间画在什么地方是无所谓的.我们也可以把上图中的虚线竖着画(图 3-2-2 右),使其原点恰好在 $P$ 处,用这个方法来标志它是 $P$ 点的切空间 $T_P$.这样切丛就有了一个形象的表示法.即令是左图,如果曲线 $\widehat{RPQ}$ 不是平面曲线,则这些"切空间"(即切线)也不会在同一个平面上,因此也不会相交.

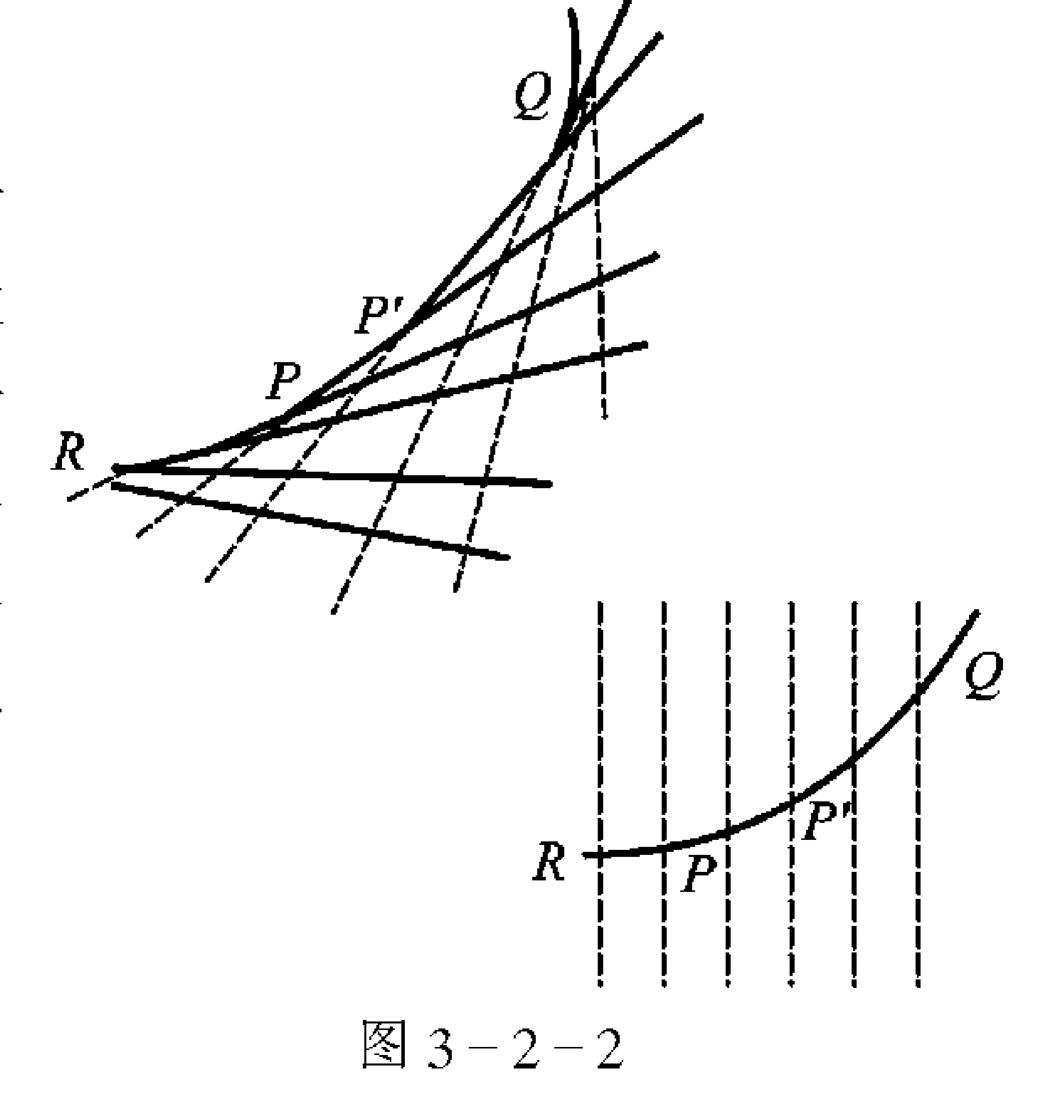


图 3-2-2

切空间、切映射和切丛在现代数学中都有特别重要的地位,我们将在第七章详细讨论.

上面我们提出了一个问题,即 $f'(x)$ 中包含了有关 $f(x)$ 的信息.现在要问,能否从 $f'(x)$ 的信息来恢复 $f(x)$? 其实这是很常见的问题:如果已知某一区间 $(a,b)$ 上的 $f'(x)$,则只要再知道 $f(x)$ 在 $(a,b)$ 之某一点 $x_0\in(a,b)$ 处之值 $f(x_0)$,则应有 $f(x)=f(x_0)+\int_{x_0}^{x} f'(t)\mathrm{d}t, x\in(a,b)$,这在物理上是非常自然的:知道了一个质点在某一时刻的位置,又知道它的速度,则此质点的任一时刻的位置自然就完全确定了.然而,我们会在第四章中看见,微分运算的逆运算究竟在什么意义下是积分运算,这决非简单问题.但是现在我们想要问的却是:如果固定了 $x$ 点,则 $f'(x)$ 能在多大程度上给出 $f$ 在此点 $x$ 处的信息? 当然,它给出了切线的斜率.但是有不少人常说,若 $f'(x)>0$,则 $f$ 在 $x$ 点单调上升.这句话是很不准确的.因为所谓"上升"、"下降"都是讲的 $f$ 在某一区间 $(a,b)$ 中的动态:若 $x_1,x_2\in(a,b)$,且由 $x_1>x_2$ 可得 $f(x_1)>f(x_2)$,则称 $f$ 在 $(a,b)$ 中上升.说 $f$ 在一点 $x$ 处上升这句话是没有意义的.我们确实可以找到这样的例子,即 $f'(0)>0$,但在包含 0 的任意小区间中,$f$ 并不上升,$f'$ 且有时正,有时负,从而 $f$ 的切线时而指向上方时而指向下方.与此相类的还有凸性的概念.例如说 $f''(x)>0$,则 $f$ 在 $x$ 点凸向下是不对的.所谓凸(这里指凸向下)也是关于 $f$ 在某一区间 $(a,b)$ 中的动态:在此区间中任取两点 $a_1$,$b_1$,如果连接 $(a_1,f(a_1))$,$(b_1,f(b_1))$ 的弦恒位于连接这两点的弧的上方,则称 $f$ 在 $(a,b)$ 中凸向下.我们这里提醒读者的是:必须注意一个函数在一点处的状态与此函数在一点的某一邻域(不论该邻域如何小)中的状态时常是两回事.读到下面的许多定理时都需要注意于此.

**2. 一般情况下微分的定义和性质** 现在考虑 $\mathbf{R}^m$ 中的开集 $U\subset\mathbf{R}^m$ 中的映射 $f: U\to\mathbf{R}^n$.我们限于考虑 $U$ 为开集是为了绕过边界点产生的困难.在 1 维情况下,开集由最多可数多个开区间 $(a,b)$ 构成.对于一个区间而言,哪怕是闭区间 $[a,b]$,其边界的构造也很简单,无非是两个点.例如对于右端 $b$ 点,我们总可以从它的左方由 $[a,b]$ 内接近于它,而可以考察其左导数.上面我们讲 $f(x)$ 的微分时就遇到过这个问题:若 $x=b$,则必须限于 $h<0$,使 $x+h=b+h$ 仍在 $[a,b]$ 内,但对高维集合 $U$,其边界构造就十分复杂了,甚至无法从 $U$ 的内部去逼近它.所以我们干脆规定 $f$ 只定义在开集 $U$ 上,这时,没有边界点在 $U$ 内,自然就不再有上述困难了.

现在在 $\mathbf{R}^m$ 中与 $\mathbf{R}^n$ 中各取坐标系 $x=(x_1,\cdots,x_m)$, $y=(y_1,\cdots,y_n)$ 则上述映射可以表示为

$$y_i=f_i(x_1,\cdots,x_m),\ i=1,2,\cdots,n. \tag{5}$$

特别是 $n=1$ 时,(5)就成了一个 $m$ 元函数

$$y=f(x_1,\cdots,x_m), \tag{6}$$

而我们将要讨论的内容就是多元函数的微分学.(5)式中的 $y=(y_1,\cdots,y_n)$ 是一个 $n$ 维向量,所以(5)就成了一个向量值函数,而时常仍用(6)式表示.

现在与定义 1 平行,我们给出

**定义 2** 设有两个有限维线性空间 $\mathbf{R}^m$ 和 $\mathbf{R}^n$ 以及 $\mathbf{R}^m$ 之开子集 $U\subset\mathbf{R}^m$,令 $f: U\to\mathbf{R}^n$ 是一个(一般为非线性的)映射,若对 $x\in U$,可以找到一个线性映射 $A(x):\mathbf{R}^m\to\mathbf{R}^n$,使得

$$f(x+h)-f(x)=A(x)h+o(h), \tag{7}$$

则称 $f$ 在 $x$ 点可微.线性映射 $A(x)$ 称为 $f$ 在 $x$ 点的微分,记作 $\mathrm{d}f(x)$:

$$A(x)=\mathrm{d}f(x). \tag{8}$$

若 $f$ 对 $U$ 之每一点均可微,则说 $f$ 在 $U$ 上可微:若 $A(x)$ 对 $x$ 连续,就说 $f(x)$ 连续可微.

现在举两个容易引起误会的例子.其一是 $f(x)$ 就是线性映射:$f(x)=Ax$.这时

$$f(x+h)-f(x)=A(x+h)-Ax=Ah.$$

而 $o(h)$ 一项变为 0.于是 $\mathrm{d}f=\mathrm{d}(Ax)=A$.我们时常就说线性变换(限于常系数)的微分即其自身,但有些文献上线性变换就用 $A$ 表示,$Ax$ 则表示此变换作用于 $x$ 所得之值.这样一来,就可能把这句话误为 $\mathrm{d}A=A$,而这个式子是很难解释的.这就涉及第二个例子,设 $f(x)=A$.这是一个常值映射:它把 $x$ 之一切值都映为同一个 $A$.这时 $f(x+h)=f(x)=A$,于是 $f(x+h)-f(x)=0$ 而有 $\mathrm{d}f=\mathrm{d}A=0$,这就把上面的 $\mathrm{d}A=A$ 弄得更糊涂了.其实,上面一个式子即古典的微积分中的 $\mathrm{d}(cx)=c\mathrm{d}x$;下面一个式子则相应于 $\mathrm{d}c=0$.这些问题都属于记号问题,或者是把映射和映射施于某个 $x$ 后所得之值混起来了.下面的定理 2 后面还有一个类似的说明,希读者注意.

还要指出,如果 $f$ 在 $x$ 可微,则其微分必是唯一的.因为若有 $A_1,A_2$ 均适合(7)式,必有

$$(A_1-A_2)h=o(h).$$

但是一个由 $\mathbf{R}^m$ 到 $\mathbf{R}^n$ 的线性映射除非是 0 映射,是不可能把任一个 $m$ 维向量 $h$ 都映为高阶无穷小量的(为什么?)所以 $A_1=A_2$ 而知 $f$ 之微分若存在必为唯一的.

定义 2 中说到了 $A(x)$ 对于 $x$ 连续.这句话的意思如下:既然 $A(x):\mathbf{R}^m\to\mathbf{R}^n$,则它必可表为一个 $n$ 行 $m$ 列矩阵 $A=(a_{ij}(x))$,$i=1,\cdots,n$,$j=1,\cdots,m$.所谓 $A(x)$ 对 $x$ 连续,现在就是指每一个 $a_{ij}(x)$ 均为 $x$ 在 $U$ 内的连续函数.当然,在更一般的理论框架下,还有进一步的说明.我们知道有限维空间一定可以赋予欧氏空间结构,即对例如 $\mathbf{R}^m$ 中一点 $x=(x_1,x_2,\cdots,x_m)$ 可以规定其范数

$$\|x\| = \Big[\sum_{i=1}^{m} x_i^2\Big]^{1/2}. \tag{9}$$

这样任意两点就有了距离,例如上述 $x$ 点和 $x'=(x_1',\cdots,x_m')$ 点的距离就规定为

$$\|x-x'\| = \Big[\sum_{i=1}^{m}(x_i-x_i')^2\Big]^{1/2}.$$

但是,$\mathbf{R}^m$ 中一点的范数可以用多种方式来定义.例如可以定义为

$$\|x\| = \sup_{1\leqslant i\leqslant m}|x_i|. \tag{10}$$

线性代数理论告诉我们,对于有限维空间,任意两种范数都是等价的.即是说,如果在 $\mathbf{R}^m$ 中有两个范数 $\|\cdot\|_1$,与 $\|\cdot\|_2$,则必存在两个常数 $c,C>0$ 使对任意一点 $x$ 均有

$$c\|x\|_2\leqslant\|x\|_1\leqslant C\|x\|_2.$$

$c,C$ 与 $x$ 无关.在下面,凡用到 $\|\cdot\|$ 时,一律是指(9)式所定义的欧几里得范数.这样,(7)式中的高阶无穷小 $o(h)$ 就理解为

$$\|o(h)\|/\|h\|\to 0 \quad (\|h\|\to 0).$$

下面考虑坐标变换下(7)式如何变化.我们先从线性空间的线性变换开始.$\mathbf{R}^m$ 和 $\mathbf{R}^n$ 中有无限多种不同的线性坐标系,如果另取

$$x'=(x_1',\cdots,x_m') \text{和} y'=(y_1',\cdots,y_n')$$

则必存在两个非奇异的线性变换

$$S=\begin{pmatrix} s_{11} & \cdots & s_{1m} \\ & & \\ s_{m1} & \cdots & s_{mm}\end{pmatrix} \quad \text{和} \quad R=\begin{pmatrix} r_{11} & \cdots & r_{1n} \\ & & \\ r_{n1} & \cdots & r_{nn}\end{pmatrix}$$

使得向量 $x=(x_1,\cdots,x_m)$ 与 $x'$ 以及 $y=(y_1,\cdots,y_n)$ 和 $y'$ 之间有

$$^t x'=S\cdot{}^t x,\ ^t y'=R\cdot{}^t y.$$

向量记号左上方的指标 $t$ 表示转置，因此例如 $^t x'$ 等都是竖向量. 这时我们有

$$^t y'=R^t y=R^t f(x)=R^t f(S^{-1}x')=(R\cdot{}^t f\cdot S^{-1})(^t x').$$

这里我们写 $^t f$ 当然是表示把 $f$（共有 $n$ 个分量）写成竖向量，其自变量最后写成 $^t x$ 也是这样. 这里没有什么问题. 若记

$$^t f'=R\,^t fS^{-1}$$

则 (7) 成为

$$f'(x'+h')-f'(x')=A'^t h'+o(h').$$

这里 $^t h'=S^{-1t}h$ 是一个 $m$ 维向量，而且与 $\|h'\|$ 与 $\|h\|$ 是同阶无穷小，因此 $o(h)$ 与 $o(h')$ 是一样的. $A'=RAS^{-1}$，所以 $A'$ 与 $A$ 是同一个对象在不同坐标系下的表现. 通常的微积分教本都只讲到一个多元函数的微分学，或者说值向量 $y$ 只是一维向量. 这时 $n=1$，从而 $R$ 是一维矩阵，即一个常数，而 $R$ 为非奇异的就意味着此常数不为 0. 不失一般性，不妨设此常数为 1. 于是 (6) 此时成为

$$y'=y=f(S^{-1}x')=(fS^{-1})(x_1',\cdots,x_m').$$

这就是一个 $m$ 元函数 $f(x)$ 在坐标变换 $x'=Sx$ 下的结果. 既然同一个映射在不同坐标系下有不同的表示，我们应考虑的是它的与坐标无关的，或称内蕴的 (coordinate - free, intrinsic) 性质. 在数学中，寻求一些数学概念或数学量的与坐标无关的性质是一件重要的工作. 例如，引进坐标是数学的一大进步，它使得许多数学概念和数学量有了具体的可以操作的形式，甚至可以计算. 只要比较一下解析几何与中学里学的初等几何学就可以体会到这一点了. 但是，引入坐标系也带来新问题，即把数学量本身的性质与一个特定坐标系的性质混为一谈了. 例如看平面上的拉普拉斯算子 $\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}$. 如果引入极坐标 $x=r\cos\theta$，$y=r\sin\theta$，经过适当的计算就会得到

$$\Delta=\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}.$$

似乎 $r=0$ 是这个算子的奇点. 但是 $r=0$ 相当于直角坐标系中的原点 (0,0)，它并不是 $\Delta=\frac{\partial^2}{\partial x^2}+\frac{\partial}{\partial y^2}$ 的奇点. 产生这个现象的原因是：原点本身就是极坐标的奇点：它的极坐标并非唯一的，而是任意的 $(0,\theta)$：$\theta$ 可以取任意值.

关于映射的微分的坐标无关的讨论见下一段. 现在我们先固定 $\mathbf{R}^m$ 和 $\mathbf{R}^n$ 的坐标系 $x$ 与 $y$，并对这个坐标系讨论映射 $f$ 可微的必要充分条件，并求出 $A(x)$ 的表达式.

**定理 1** 定义 2 中的映射 $f$ 在某坐标系下在开集 $\Omega$ 中连续可微的必要充分条件是 $f$ 的各个分量 $y_i$ 对任意 $x_j$ 均为连续可微函数. 这时

$$A(x)=\frac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_m)}=\begin{pmatrix}\frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_m}\\ \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_m}\end{pmatrix}. \tag{11}$$

(11)右方矩阵称为雅可比矩阵(Jacobian matrix).

**证** 先证充分性.我们不妨只看 $n=1$, $m=2$ 的特例.这时 $h=(h_1,h_2)$ 而且映射 $f$ 就是 $f=f_1$,于是设它是一个连续可微函数:

$$\begin{aligned} f(x+h)-f(x) &= f(x_1+h_1,x_2+h_2)-f(x_1,x_2) \\ &= [f(x_1+h_1,x_2+h_2)-f(x_1,x_2+h_2)]+ \\ &\quad [f(x_1,x_2+h_2)-f(x_1,x_2)]. \end{aligned} \tag{12}$$

当 $f$ 对 $(x_1,x_2)$ 在开集 $\Omega$ 中均为连续可微时,对右方两项分别应用拉格朗日公式有

$$f(x+h)-f(x)=\frac{\partial f}{\partial x_1}(x_1+\theta_1 h_1,x_2+h_2)h_1+\frac{\partial f}{\partial x_2}(x_1,x_2+\theta_2 h_2)h_2.$$

再利用 $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$ 在 $(x_1,x_2)$ 处的连续性即得

$$\begin{aligned} f(x+h)-f(x) &= \left(\frac{\partial f}{\partial x_1}(x)h_1+\frac{\partial f}{\partial x_2}(x)h_2\right)+o(h) \\ &= A(x)h+o(h). \end{aligned}$$

而这里 $A(x)$ 是 $1\times 2$ 矩阵 $\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2}\right)$.由假设,它是连续的,于是映射 $f$ 在此开集中连续可微,而且其微分就是雅可比矩阵.充分性得证.

必要性很简单.设 $f$ 在此开集中连续可微,任取此开集中一点 $x$,对 $f$ 之任一分量(不妨即设为 $f_1=y_1$),对 $h=(h_1,0)$ 有

$$f(x_1+h_1,x_2)-f(x_1,x_2)=A(x)h_1+o(h).$$

双方除以 $h_1$ 再令 $h_1\to 0$,即知 $\frac{\partial f}{\partial x_1}$ 连续.同理 $f$ 之各分量对各个 $x_j$ 之偏导数 $\frac{\partial y_i}{\partial x_j}$ 均在此开集中连续而必要性得证.

**注 1** 在 $U$ 中连续可微的映射 $f$ 之集合构成一个空间.我们通常记作 $C^1(U)$.上面说到了这时 $\mathrm{d}f=A(x)$ 是雅可比矩阵,而且其元素都是古典意义下的连续可微函数.这类函数之集合也记作 $C^1(U)$.这样看来,$C^1(U)$ 的两种说法是一致的,我们不必加以区别.不过,后一种意义下说 $A(x)$ 之元是 $C^1$ 函数是在选定了一个坐标系以后说的.如果变一个坐标系,它是否仍在 $C^1(U)$ 中?答案是肯定的,下面会详细解释.

**注 2** 定义 2 中讲到了 $f$ 在一点 $x$ 可微.因此有一个错觉,以为其必要充分条件应是 $\frac{\partial y_i}{\partial x_j}$ 在 $x$ 点存在.实际上不是这样,而只当 $m=1$ 时它是对的,因为 $m=1$ 时,1 维向量 $h$ 就是一个数 $h$,向量 $h\neq 0$ 就是数 $h\neq 0$.于是对在定义 1 的(4)式(那里 $n$ 也是 1,实际上 $n>1$ 也对)双方可以用 $h$ 去除,而当 $h\to 0$ 时立即有 $A=\frac{\mathrm{d}f(x)}{\mathrm{d}x}=f'(x)$. $m>1$ 时我们就不能用 $m$ 维向量 $h$ 去除(7)式两侧,而要利用(12)式过渡到一个自变量的情况,并且应用拉格朗日公式.这就要用到 $x$ 点以外其它点处偏导数的存在.然后还要利用其在 $x$ 点的连续性,把这些偏导数化为在 $x$ 点的值再加上一个无穷小量.这就看到,$m>1$ 时可微性与可求导数是不相同的,二者有本质区别.但是我们无法把可微性准确地刻画为可求导,再加上导数应适合什么条件,所以,我们干脆讨论某一开集中

的连续可微性，而把我们需要的结果以定理 1 的形式陈述出来. 这样做不但是够用了，而且还想借此再次提醒读者：一个函数（即一个映射）在一点的情况与在一点的某个邻域中的情况是不一样的. 当我们讨论一个函数的可微性时，真正重要的是此函数在一点邻域中的情况. 再从另一方面说，若我们确定了一个坐标系 $(x_1,\cdots,x_m)$，则点 $x$ 的变化可以是沿某个坐标轴，例如 $x_1$ 轴. 所以 $h=(h_1,0,\cdots,0)$，而我们可以写出

$$\Delta_1 f = f(x_1,+h_1,x_2,\cdots,x_m)-f(x)=A_1(x)h_1+o(h_1), \tag{13}$$

这样的 $A_1(x)$ 相应于偏导数，或称为偏微分. 为了 $f$ 有偏微分，由极限式

$$\lim_{h_1\to 0}\frac{\Delta_1 f}{\Delta x_1}=A_1(x),\ \Delta x_1=h_1$$

可知只要 $f$ 为可求导即可. 与偏微分相对立，(7) 中的 $A(x)$ 称为全微分. 偏微分亦称加托 (Gateaux) 微分，全微分则称弗雷歇 (Fréchet) 微分. 二者的关系我们就不讲了. 但由此容易理解为什么 $m=1$ 与 $m>1$ 时可微性的情况不一样.

**注 3** 现在，微分是一个矩阵. 线性代数告诉我们一个 $n\times m$ 矩阵 $A$，就是一个 $\mathbf{R}^m$ 到 $\mathbf{R}^n$ 的线性映射. 我们把映射 $f:\Omega\to\mathbf{R}^n$ 的 $\mathrm{d}f$ 微分也定义为一个映射. 当 $m=n=1$ 时，我们会有一个一阶矩阵，但是一阶矩阵也就是一个数. 所以这时 $A(x)=(\mathrm{d}f)(x)$ 按古典的微分的观点看来是一个数，即导数 $f'(x)$. 但是按我们现在的观点看来则是一个映射 $\mathrm{d}f:\mathbf{R}^1\to\mathbf{R}^1$. 按古典的观点 $\mathrm{d}f=f'(x)\mathrm{d}x=f'h$，按现在的观点则 $\mathrm{d}f=f'(x)$，所以在一些文献中 $\mathrm{d}f$ 就表示 $f'(x)$. 这种记号上的混淆只有请读者自己注意. 倒是值得注意的是，$\mathrm{d}f$ 在固定的坐标系下要用雅可比矩阵

$$J(x)=\frac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_m)}=\begin{pmatrix}\dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \cdots & \dfrac{\partial y_1}{\partial x_m}, \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_2}{\partial x_m} \\ & & & \\ \dfrac{\partial y_n}{\partial x_1} & \dfrac{\partial y_n}{\partial x_2} & \cdots & \dfrac{\partial y_n}{\partial x_m}\end{pmatrix} \tag{14}$$

表示，于是也就可以说雅可比矩阵是 $f'(x)$ 的合适的推广.

还有一点要提醒，在古典的微积分教本中是在 $m=n$ 时讲雅可比行列式，并且采用下面的记号

$$J(x)=\frac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_n)}=\begin{vmatrix}\dfrac{\partial y_1}{\partial x_1} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ & & \\ \dfrac{\partial y_n}{\partial x_1} & \cdots & \dfrac{\partial y_n}{\partial x_n}\end{vmatrix}. \tag{15}$$

但 $m\neq n$ 时就谈不到雅可比行列式，所以我们只讲雅可比矩阵，$J(x)$ 与 $\dfrac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_m)}$ 则按 (14) 理解，而把雅可比行列式写作 $\det J(x)$ 或 $\det\dfrac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_m)}$. 这些也要读者多加小心.

关于为什么雅可比矩阵是 $f'(x)$ 的合适的推广，下面还要详细讲.

**注 4** 上面我们讲到在 $m>1$ 时，必须利用(12)式过渡到一个自变量的情况，然后再用拉格朗日公式. 于是要问，在 $n>1$ 时是否也有与拉格朗日公式相应的公式？很遗憾，没有. 而且有反例说明不可能有. 但是注意到即使在一元函数情况下，公式 $f(b)-f(a)=f'(\xi)(b-a)$ 中的 $\xi$ 之确切位置也并不清楚，而在使用上式时，我们时常只能满足于一个不等式，即由 $m\leqslant|f'(\xi)|\leqslant M$ 可以得出

$$m|b-a|\leqslant|f(b)-f(a)|\leqslant M|b-a|.$$

在 $n>1$ 时我们也可以得到这样一个不等式. 而且我们就把相应的不等式称为 $n>1$ 时的中值定理：

**定理 2（中值定理）** 设在一维空间 $\mathbf{R}$ 中有闭区间 $I=[a,b]$. $f: I\to\mathbf{R}^n$ $(n\geqslant 1)$ 在 $[a,b]$ 上连续而在开区间 $(a,b)$ 上可微，则若

$$\|A(x)\|\leqslant M, \qquad x\in(a,b), \tag{16}$$

必有

$$\|f(b)-f(a)\|\leqslant M(b-a). \tag{17}$$

**证** 先解释一下记号. 现在我们处理的情况是 $m=1, n\geqslant 1$. 因此，$A(x)$ 是一个 $n\times 1$ 矩阵即一个竖向量. $f(x)$ 也是一样. 竖向量的范数和横向量的范数一样，我们可以取为欧几里得范数. 但是我们想借此机会讲一下一般的线性算子 $A(x): \mathbf{R}^m\to\mathbf{R}^n$ 的范数如何定义. $A(x)$ 作用于 $h\in\mathbf{R}^m$，我们定义算子 $A(x)$ 之范数为

$$\|A(x)\|=\sup\frac{\|A(x)h\|}{\|h\|}. \tag{18}$$

这个定义对于更一般的空间(包括无穷维空间)上的线性算子也是适用的. 但现在 $A(x)$ 是一个 $n\times m$ 矩阵，我们自然要问 $\|A(x)\|$ 怎样用 $A(x)$ 的元素来表示呢？这就要看如何定义 $\|h\|$ (虽然 $\|h\|$ 之不同定义是等价的，但却会影响 $\|A(x)\|$ 之定义). 这个问题我们在此不讲了. 现在只提到 $\|A(x)\|$ 以 $x$ 为参数. 若当 $x$ 在一个有界闭集(以后将说明这种集合应称为紧集，详见第六章)中变化时 $A(x)$ 之元素是连续的，则 $\|A(x)\|$ 有界：$\sup\limits_x\|A(x)\|<+\infty$，所以，(17)中的 $M$ 就是 $\sup\limits_x\|A(x)\|$.

这个定理看起来很简单，其实并不容易证. 我们并未假设 $f(x)$ 在区间端点 $a, b$ 可微. 那么，拉格朗日公式中也未假设 $f(x)$ 在 $x=a, b$ 时有导数，为什么又很容易证呢？这是因为 $n=1$ 时有罗尔定理可用，而 $n>1$ 时就没有了. 罗尔定理是一个不起眼的定理. 罗尔其人也是一个不起眼的人. 其实，此人在微分学发展中有过重要贡献，他的定理也应该注意.

定理证明如下，先证对任意小的正数 $\eta>0$ 均有

$$\|f(b)-f(a)\|\leqslant(M+\eta)(b-a)+\eta. \tag{19}$$

然后再令 $\eta\to 0$ 即得(17). (19)中右方有一个多余的 $\eta$ 在，这是由于我们只假设了 $f(x)$ 在 $(a,b)$ 中可微，而不是在 $[a,b]$ 上可微. 这一点从证明过程即可看到.

(19)的证明表面上很简单，读者可能会想，既已假设 $f$ 可微，所以

$$\begin{aligned}\|f(a+h)-f(a)\| &= \|f'(a)h+o(h)\|\\ &\leqslant(M+\eta)\|h\|+\eta.\end{aligned}$$

实际上远不如想象的简单. 一则 $f$ 在 $a$ 点并不可微. 因此上式第一行并不一定成立. 其次，也更重要的是尽管 $o(h)$ 对 $h$ 是高阶无穷小，但不等式 $\|o(h)\|\leqslant\eta\|h\|$ 是当 $\|h\|$ 很小时才成立的，

现在 $h=b-a$，所以 $\|h\|=b-a$ 并不很小，从而 $\|o(h)\|\leqslant\eta\|h\|$ 也不一定对.因此，我们要采用一种所谓“连续拓展法”来证明它，即先证明在 $x=a$ 附近有

$$\|f(x)-f(a)\|\leqslant(M+\eta)(x-a)+\eta.$$

因此这个不等式可以从 $a$ 向右推开去.推到什么地方为止？如果推到某一个子区间，标准的方法是证明这个子区间在 $[a,b]$ 中又开又闭，而 $[a,b]$ 中又开又闭又非空的子区间就是 $[a,b]$.由此定理得证.但是这样就必须用拓扑学的语言，而我们并不假设读者具有这种准备.所以下面我们用的思想来自拓扑学，但语言来自古典的微积分.

考虑一个集合

$$J=\{\xi\in[a,b]，当\ a\leqslant x\leqslant\xi\ 时，$$
$$\|f(\xi)-f(a)\|\leqslant(M+\eta)(\xi-a)+\eta\}.$$

先证明 $J$ 是非空的.事实上，当 $\xi=a$ 时，上式当然成立.因为 $f(x)$ 在 $a$ 连续，必有一个正数 $\rho$，使当 $a\leqslant x\leqslant a+\rho$ 时

$$\|f(x)-f(a)\|\leqslant\eta\leqslant(M+\eta)(x-a)+\eta.$$

注意，我们这里只应用了 $f(x)$ 在 $a$ 连续.而完全不需要 $f(x)$ 在 $a$ 可微.总之，$[a,a+\rho]$ 在 $J$ 中.我们令

$$c=\sup J,$$

于是有一串 $c_n\to c$，而

$$\|f(c_n)-f(a)\|\leqslant(M+\eta)(c_n-a)+\eta.$$

令 $c_n\to c$ 即有

$$\|f(c)-f(a)\|\leqslant(M+\eta)(c-a)+\eta.$$

所以 $c\in J$.现在有两个可能性，首先是 $c=b$，这就是说 $[a,b]$ 全在 $J$ 中，从而在 $[a,b]$ 上(19)成立.在(19)式中令 $\eta\to 0$ 即得定理之证.或者是 $c<b$，这时 $a<c<b$，而 $f(x)$ 在 $c$ 可微.我们终于可以利用 $f(x)$ 在 $x=c$ 处可微了.

取一个正数 $\delta$ 使 $c+\delta\leqslant b$，只要 $\delta$ 充分小，对于 $z\in(c,c+\delta)$ 有

$$f(z)-f(c)=f'(c)(z-c)+o(z-c).$$

当 $\delta$ 充分小时，$|z-c|<\delta$，从而 $\|o(z-c)\|<\eta(z-c)$，又因 $\|f'(c)\|\leqslant M$，所以

$$\|f(z)-f(c)\|\leqslant(M+\eta)(z-c).$$

因此

$$\begin{aligned}\|f(z)-f(a)\|&\leqslant\|f(c)-f(a)\|+\|f(z)-f(c)\|\\&\leqslant(M+\eta)(c-a)+(M+\eta)(z-c)+\eta\\&\leqslant(M+\eta)(z-a)+\eta.\end{aligned}$$

所以 $c+\delta\in J$.这与 $c=\sup J$ 矛盾.于是定理得证.

作为中值定理的一个应用，我们来证明一个十分重要而证明起来颇不简单的、似乎不言而喻的定理.

**定理 3** 设 $f: U\subset\mathbf{R}^m\to\mathbf{R}^n$ 是定义 2 中讲的可微函数.若 $A(x)=0$ 于 $U$ 中，则 $f(x)\equiv\text{con st}$.这里设 $U$ 是连通的.

**证** 任取一点 $A\in U$，设其坐标为 $x_A$，因为 $U$ 是开集所以必有一以 $A$ 为心的球含于 $U$ 内.今证在此球内 $f(x)=f(x_A)$.为此，在球内任取一点 $B$，设其坐标为 $x_B$，于是联结 $AB$ 的线段全

在球内，其方程为

$$x=(1-t)x_A+tx_B, 0\leqslant t\leqslant 1.$$

$A$ 点与 $B$ 点分别对应于参数值 $t=0$，$t=1$. 令

$$\Phi(t)=f[(1-t)x_A+tx_B].$$

于是 $\Phi(t): I\to\mathbf{R}^n$，$I=[0,1]$. 很明显 $\Phi(t)$ 是 $t$ 的可微函数. 因为已设 $A(x)=0$，所以 $\Phi'(t)=0$，$t\in I$，而可取(16)中的 $M=0$. 对于 $\Phi(t)$ 显然可适用中值定理，从而

$$\Phi(1)-\Phi(0)=0.$$

或 $f(x_B)=f(x_A)$. 此式在上述球内之任一点 $x_B$ 处均成立，所以在此球内有 $f(x)\equiv f(x_A)$.

现在可以用此球内任一点作为心，再作其它的球，只要此球含于 $U$ 内，必有 $f(x)\equiv f(x_A)$ 于其中. 现在令

$$V=\{x; x\in U, f(x)=f(x_A)\}.$$

于是 $V$ 非空. $V$ 一定是开的，因为若 $x_0\in V\subset U$，$x_0$ 必为开集 $U$ 之内点，因此必有以 $x_0$ 为心的球 $W\subset U$，而且由上证，在 $W$ 内 $f(x)=f(x_0)=f(x_A)$，而 $W\subset V$. 这样，$V$ 一定是非空开集. 另一方面 $V$ 又一定是闭的. 因为若 $x_0\in\overline{V}$ 而且 $x_0\in U$，则由定理的条件，$f(x)$ 在 $x_0$ 为连续的. $x_0$ 在 $V$ 的闭包中，所以存在一个序列 $\{x_n\}\subset V$，$x_n\to x_0$. 由 $f(x)$ 之连续性，$f(x_0)=\lim\limits_{n\to\infty}f(x_n)=\lim\limits_{n\to\infty}f(x_A)=f(x_A)$. 因此 $x_0\in V$，这样又得到 $V$ 是闭的. 一个连通集 $U$ 的非空的既开又闭的子集必是 $U$ 本身，所以 $V=U$. 即在 $U$ 中 $f(x)=f(x_A)$. 定理证毕.

**注 1** 这个定理的证明有两点特别值得注意. 首先是我们先把待证的定理化到一个球内. 球的特点是：球内任一点均可用直线段(实即半径)与球心连接. 凡具有这种性质的区域称为对一点(现在的球心)为星形的区域. 而在直线段上，自变量 $t$ 恒在 $I=[0,1]$ 中，而问题化为 $m=1$，$n\geqslant 1$ 的特例. 这是一个常用的方法，下面讲泰勒公式时还要用它，这是一个与拓扑学的一些重要概念有关的方法.

**注 2** 我们两次使用了连续拓展法. 第一次读者可能不感到太奇怪，第二次则可能有不少读者感到生疏. 可能人们会以为，既已证得在一球内 $f(x)\equiv f(x_A)$，仿此可以作一串球，一直达到 $x_B$ 处，然后即知在 $U$ 内 $f(x)\equiv f(x_A)$. 但是稍想一下，真正可以作一串球从 $x_A$ 一直拓展到 $x_B$ 吗？直觉地来看确是如此，但是考虑到高维空间区域的构造可能极为复杂，这时读者就不会那么有把握了. 我们第二次使用这个方法倒不止是拓扑学的思想，微积分的语言了，而是拓扑学的证明. 读者一定会问，何以连通性就会得出这一大套结论？读到第六章时才回过来看看这个证明就容易懂了.

**注 3** 这个定理假设了 $U$ 是连通的. 若没有这个假设，$U$ 必可分成若干个连通子区域之并：$U=\bigcup\limits_i U_i$，每一个 $U_i$ 都是连通的. 若在不同的 $U_i$ 中各取 $x_A$，当可证明 $f(x)$ 在每个 $U_i$ 中分别等于一个常数：$f(x)=C_i$，$x\in U_i$，而这些 $C_i$ 彼此可能不同. 这种 $f(x)$ 称为局部常值函数.

下面我们来看一下微分作为一个运算，亦即微分算子 d 的性质.

第一个性质几乎是自明的，即**线性性质**

$$\mathrm{d}(C_1f_1+C_2f_2)=C_1\mathrm{d}f_1+C_2\mathrm{d}f_2, C_1, C_2 \text{ 是常数}. \tag{20}$$

第二个性质称为**莱布尼茨性质**：若 $f_1, f_2$ 可微，则 $f_1f_2$ 也可微，而且以下的莱布尼茨公式成

立：

$$d(f_1 f_2) = f_1 df_2 + f_2 df_1. \tag{21}$$

证明很简单：

$$\begin{aligned}&(f_1 f_2)(x+h) - (f_1 f_2)(x)\\ &= [f_1(x+h) - f_1(x)]f_2(x+h) + f_1(x)[f_2(x+h) - f_2(x)]\\ &= [df_1 \cdot h + o(h)][f_2(x) + o(1)] + f_1(x)[df_2 \cdot h + o(h)]\\ &= (f_2 df_1 + f_1 df_2)h + o(h).\end{aligned}$$

在数学中有许多乘积：向量的数量积、向量积、外积等等，也有许多求导或微分：普通的求导和微分、外微分、协变导数等等，而一定要适合(21)(或者其某一项前加上一个因子 $(-1)^s$)这样的公式.所以，只要有这种类型的公式成立，我们就称相应的算子为导子(derivation)或反导子(antiderivation)即 $(-1)^s = -1$ 的情况.

可是关于微分算子 d 最重要的性质无疑是下面的链式法则，而这就导致对微分作内蕴的处理.

**3. 微分的内蕴性质** 以上的讨论都是对某一固定的坐标系进行的，所以要问如果有了坐标变换，这些结果是否仍成立，或有改变？如果有改变，则新老坐标系中相应的结果有什么关系？为了回答这个问题，首先要问，哪一些坐标变换是容许的.这就导致了微分同胚的概念.

**定义 3** 设有 $\mathbf{R}^n$ 中的开区域 $U$ 与 $V$，各赋有坐标 $x=(x_1,\cdots,x_n)$ 与 $y=(y_1,\cdots,y_n)$.若 $U$ 与 $V$ 之间有一映射 $\varphi: U\to V$，$y_i=\varphi_i(x_1,\cdots,x_n)$，它有逆映射 $\psi=\varphi^{-1}: V\to U$，$x_i=\psi_i(y_i,\cdots,y_n)$，从而实现了 $U$ 与 $V$ 之间的一一对应；再设 $\varphi_i$ 与 $\psi_i$ 分别在 $U$ 与 $V$ 中为 $C^k$ 可微，则称 $\varphi$ (或 $\psi$)为 $U\to V$ (或 $V\to U$)的 $C^k$ 微分同胚.

现在来看链式法则.我们有

**定理 4** 设 $U\subset\mathbf{R}^m$ 是一个开集，$x_0\in U$. $f: U\to\mathbf{R}^n$ 在 $x_0$ 处可微.又设 $g: V\to\mathbf{R}^p$ 在 $y_0=f(x_0)$ 处可微，这里 $V\subset\mathbf{R}^n$ 且 $f(U)\subset V$.则 $\Phi=g\circ f: U\to\mathbf{R}^p$ 在 $x_0$ 处也可微，而且

$$d\Phi(x_0) = dg(y_0)\circ df(x_0). \tag{22}$$

**证** 在证明之前先要回顾一下可微映射的定义.关于映射 $f: U\to\mathbf{R}^n$ 的可微性的定义中，在 $\mathbf{R}^m(U\subset\mathbf{R}^m)$ 与 $\mathbf{R}^n$ 中都是取定了坐标系的.所以我们先在 $\mathbf{R}^m$，$\mathbf{R}^n$ 与 $\mathbf{R}^p$ 中都取定坐标系，这样 $df(x_0)$ 与 $dg(y_0)$ 是两个确定的矩阵.现在由可微性的定义即有

$$f(x_0+h) = f(x_0) + df(x_0)\cdot h + o_1(h),$$

$$g(y_0+k) = g(y_0) + dg(y_0)\cdot k + o_2(k),$$

而且对任意 $\varepsilon_1>0$，必可找到 $\delta_1>0$ 使当 $\|h\|<\delta_1$ 时，

$$\|o_1(h)\| < \varepsilon_1\|h\|,$$

对于 $k$ 也有 $\varepsilon_2,\delta_2$，使当 $\|k\|<\delta_2$ 时，$\|o_2(k)\|<\varepsilon_2\|k\|$.而且因为 $df(x_0)$，$dg(y_0)$ 是确定的矩阵，所以必有适当的正数 $M>0$ 使

$$\|df(x_0)\cdot h\| \leqslant M\|h\|, \quad \|dg(y_0)\cdot k\| \leqslant M\|k\|.$$

现在我们来计算 $\Phi(x_0+h)-\Phi(x_0)$.我们有

$$\Phi(x_0+h) = (g\circ f)(x_0+h$$

$$= g(y_0+k) = g(y_0) + \mathrm{d}g(y_0)k + o_2(k),$$

这里 $k = \mathrm{d}f(x_0)h + o_1(h)$. 所以当 $\|h\| < \delta$ 时,

$$\|k\| \leqslant \|\mathrm{d}f(x_0)h\| + \|o_1(h)\| \leqslant (M+\varepsilon_1)\|h\|$$
$$\leqslant (M+1)\|h\|.$$

所以

$$\Phi(x_0+h) = \Phi(x_0) + \mathrm{d}g(y_0)[\mathrm{d}f(x_0)h + o_1(h)] + o_2(k)$$
$$= \Phi(x_0) + \mathrm{d}g(y_0)\cdot\mathrm{d}f(x_0)h + \mathrm{d}g(y_0)o_1(h) + o_2(k). \quad (23)$$

当 $\|h\| < \min\left(\delta_1, \dfrac{\delta_2}{M+2}\right)$时,一方面有

$$\|\mathrm{d}g(y_0)o_1(h)\| \leqslant M\varepsilon_1\|h\|, \quad (24)$$

同时又有

$$\|k\| \leqslant (M+1)\|h\| < \delta_2,$$

从而

$$\|o_2(k)\| < \varepsilon_2\|k\| \leqslant (M+1)\varepsilon_2\|h\|. \quad (25)$$

合并(24),(25)即知当 $\|h\| < \min\left(\delta_1, \dfrac{\delta_2}{M+2}\right)$时

$$\|\mathrm{d}g(y_0)o_1(h) + o_2(k)\| < (M\varepsilon_1 + (M+1)\varepsilon_2)\|h\|.$$

由 $\varepsilon_1, \varepsilon_2$ 任意性,知道上式右方比之 $h$ 是高阶无穷小. 所以(23)式最终化为

$$\Phi(x_0+h) = \Phi(x_0) + \mathrm{d}g(y_0)\cdot\mathrm{d}f(x_0)h + o(h),$$

因此 $\Phi$ 在 $x_0$ 是可微的,而且其微分是

$$\mathrm{d}\Phi(x_0) = \mathrm{d}g(y_0)\circ\mathrm{d}f(x_0).$$

这样在指定的坐标系下定理得到了证明. 当然,对任意的坐标系这个证明均有效,所以(22)式成立. 证明暂停于此.

余下的是,如果在 $U$ 中作坐标变换

$$x_i' = X_i(x_i, \cdots, x_m),\ i = 1,2,\cdots,m.$$

类似地,也在 $\mathbf{R}^n$ 和 $\mathbf{R}^p$ 中作坐标变换

$$y_i' = Y_i(y_1, \cdots, y_n),\ i = 1,2,\cdots,n,$$
$$z_i' = Z_i(z_1, \cdots, z_p),\ i = 1,2,\cdots,p,$$

会得到什么结果? 这里有两个问题,首先,上述坐标变换是 $C^k$ 微分同胚,而且因为问题只涉及一阶导数,所以只需考虑 $k=1$ 的情况. 这样的微分同胚是否存在? 这是一个很大的问题. 下面我们会看到,局部的微分同胚总是存在的,而整体的则不一定. 但是迄今为止,我们引进的概念全是局部的,所以不妨设 $U$ 是 $x_0$ 的一个充分小的邻域,它在微分同胚下变成 $x'_0$ 的邻域 $U'$. 对于 $y_0 = f(x_0)$,我们不必设 $g: \mathbf{R}^n \to \mathbf{R}^p$,而只要设 $g: V \to \mathbf{R}^p$ 为可微. $V$ 是 $y_0$ 的一个邻域,而且 $f(U) \subset V$, $y_0$ 与 $V$ 在微分同胚 $Y$ 下变为 $y_0'$ 与 $V'$. 最后,令 $z_0 = g(y_0) = \Phi(x_0)$, $z_0$ 有一个邻域 $W$,使 $g(V) \subset W$. 同样,在微分同胚 $Z$ 下也有 $z_0'$ 与 $W'$. 其次,在不同坐标系下得出的微分如 $\mathrm{d}\Phi$, $\mathrm{d}g$ 与 $\mathrm{d}f$ 有什么关系. 这里我们先统一一下我们的用语. 如果 $f: U \to V$ 是一个映射,我们称 $U$ 为映射的源空间(source space), $V$ 为映射的靶空间(target space). 于是我们的问题成为,若在

源空间（靶空间）中作坐标变换，$\mathrm{d}f$ 将如何变换？为了回答这个问题我们需要一个引理

**引理 5** 若 $X: U \to U'$ 是一微分同胚，则 $\mathrm{d}X \cdot \mathrm{d}X^{-1} = I$，$\mathrm{d}X^{-1} \cdot \mathrm{d}X = I$.

**证** $X$，$X^{-1}$ 以及 $id$（恒等映射）都是可微的，所以可以应用定理 4 而有

$$\mathrm{d}(id) = \mathrm{d}X \circ \mathrm{d}X^{-1}, \mathrm{d}(id) = \mathrm{d}X^{-1} \circ \mathrm{d}X.$$

但是 $id$ 是一个常系数的线性变换，定义 2 后紧接着就说了，其微分等于其自身，即也是一个恒等变换 $I$，所以引理得证.

我们不妨在一个坐标系下看看这个引理意味着什么. 定理 1 指出，微分就是雅可比矩阵：

$$\mathrm{d}X = \frac{\partial(x_1', \cdots, x_n')}{\partial(x_1, \cdots, x_n)}, \ \mathrm{d}X^{-1} = \frac{\partial(x_1, \cdots, x_n)}{\partial(x_1', \cdots, x_n')}.$$

所以，引理 5 用矩阵来表达就是

$$\frac{\partial(x_1', \cdots, x_n')}{\partial(x_1, \cdots, \partial x_n)} = \left[\frac{\partial(x_1, \cdots, x_n)}{\partial(x_1', \cdots, x_n')}\right]^{-1}.$$

这个结论用矩阵乘法直接验证并不难. 因为这两个矩阵之积 $\frac{\partial(x_1', \cdots, x_n')}{\partial(x_1, \cdots, x_n)} \cdot \frac{\partial(x_1, \cdots, x_n)}{\partial(x_1', \cdots, x_n')}$ 之第 $k$ 行第 $l$ 列的元是

$$\sum_m \frac{\partial x_k'}{\partial x_m} \frac{\partial x_m}{\partial x_l'} = \frac{\partial x_k'}{\partial x_l'} = \delta_{kl}.$$

但是引理 5 的表述显然更能说明问题的本质.

读者又可能会问，上面的记号 $\mathrm{d}X^{-1}$ 究竟是指逆映射 $X^{-1}$ 的微分 $\mathrm{d}(X^{-1})$，还是微分 $\mathrm{d}X$ 作为一个映射的逆映射 $(\mathrm{d}X)^{-1}$？其实，引理 5 正是说明，它们是一回事，而且既是左逆又是右逆.

现在来看源空间中的坐标变换 $X: U \to \tilde{U}$. 原来的映射 $f: U \to V$ 衍生出新映射 $\tilde{f}: \tilde{U} \to V$，这里 $\tilde{f} = f \circ X^{-1}$. 于是由定理 4 和引理 5 有

$$\mathrm{d}\tilde{f} = \mathrm{d}f \circ \mathrm{d}X^{-1}.$$

因此，源空间的坐标变换相应于将微分 $\mathrm{d}f$ 右乘以 $\mathrm{d}X^{-1}$. 类似于此，若在靶空间中作坐标变换 $Y: V \to \tilde{V}$，原来的映射将衍生出新映射 $\tilde{f}: U \to \tilde{V}$，$\tilde{f} = Y \circ f$. 于是又有

$$\mathrm{d}\tilde{f} = \mathrm{d}Y \circ \mathrm{d}f.$$

所以靶空间中的坐标变换相当于用 $\mathrm{d}Y$ 左乘 $\mathrm{d}f$. 所以，如果对定理 4 中的 $U$，$V$，$W$ 都作变换 $X$，$Y$，$Z$，则 $f$ 生成

$$\tilde{\Phi} = Z \circ g \circ Y^{-1} \circ Y \circ f \circ X^{-1} = \tilde{g} \circ \tilde{f},$$

$$\mathrm{d}\tilde{\Phi} = \mathrm{d}Z \circ \mathrm{d}g \circ \mathrm{d}Y^{-1} \circ \mathrm{d}Y \circ \mathrm{d}f \circ \mathrm{d}X^{-1} = \mathrm{d}\tilde{g} \circ \mathrm{d}\tilde{f}. \tag{26}$$

不过这里的 $\tilde{f}$ 与上面的不同：在上面 $\tilde{f}$ 是由 $f$ 经过源空间的坐标变换而得，故 $\tilde{f} = f \circ X^{-1}$，现在则源空间与靶空间都作了变换，所以 $\tilde{f} = Y \circ f \circ X^{-1}$. 同样，$\tilde{g} = Z \circ g \circ Y^{-1}$. (26) 形式上与 (22) 是一样的. 这就是说，无论坐标系如何选取，(22) 都是成立的. 也就是说 (22) 是一个与坐标选择无关的结果，定理 4 证明至此完成.

定理 4 是一个很深刻的结果. 它告诉我们，两个映射的复合原本是一个很复杂的映射. 但是如果考虑其线性化，则 $\mathrm{d}\Phi$ 等于两个线性映射的复合：$\mathrm{d}\Phi = \mathrm{d}g \circ \mathrm{d}f$. 线性映射可以用一个矩阵来

表现.它的复合很简单:先用一个矩阵作用于一个向量(例如前述的 $h$),再用第二个矩阵作用上去.本来,$\Phi=g\circ f$ 的意思也是先对源空间中的元施以 $f$,再继续施以 $g$.不过 $g\circ f$ 到底是什么就很难说了,$f$ 和 $g$ 的定义域和值域的关系也会出麻烦.现在好了,反正 $\mathrm{d}g$ 和 $\mathrm{d}f$ 都是线性变换,其定义域总是一个整个的线性空间,例如 $\mathrm{d}f$ 定义在 $\mathbf{R}^m$ 上.尽管 $\mathrm{d}f$ 不一定把整个 $\mathbf{R}^m$ 映到整个 $\mathbf{R}^n$ 上,但其值域总在 $\mathbf{R}^n$ 中,即在 $\mathrm{d}g$ 的定义域内,因此 $\mathrm{d}g\circ\mathrm{d}f$ 总是有意义的.一般的映射就不一定能行,所以上面才有例如 $f(U)\subset V$, $g(V)\subset W$ 这样的限制.其次是 $\mathrm{d}g\circ\mathrm{d}f$ 在一定坐标系下很容易计算,它就是矩阵的乘法.这是因为,若我们把 $f$ 写成 $y_i=f_i(x_1,\cdots,x_m)$, $i=1,\cdots,n$,则 $\mathrm{d}f$ 的矩阵表示是 $\dfrac{\partial(f_1,\cdots,f_n)}{\partial(x_1,\cdots,x_m)}=\dfrac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_m)}$.同样,若把 $g$ 写成 $w_i=g_i(y_1,\cdots,y_n)$,则 $\mathrm{d}g$ 的矩阵表示是 $\dfrac{\partial(g_1,\cdots,g_p)}{\partial(y_1,\cdots,y_n)}$,而且若以 $y_i=f_i(x)$ 代入 $g_i$ 后,$w_i$ 将是 $\Phi(x)$ 的坐标:$w_i=\Phi_i$.但是 $\dfrac{\partial(g_1,\cdots,g_p)}{\partial(y_1,\cdots,y_n)}\dfrac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_m)}$ 的第 $k$ 行 $l$ 列的元是

$$\sum_{i=1}^{n}\frac{\partial g_k}{\partial y_i}\frac{\partial y_i}{\partial x_l}=\frac{\partial g_k}{\partial x_l}.$$

所以上述矩阵之积是 $\dfrac{\partial(w_1,\cdots,w_p)}{\partial(x_1,\cdots,x_m)}$ 即 $\mathrm{d}\Phi$ 的矩阵表示.前面我们已经说了,$\mathrm{d}X^{-1}$ 的矩阵表示就是 $\mathrm{d}X$ 的矩阵表示的逆矩阵,现在又看到线性变换之积(即复合)对应于相应矩阵的乘积.所以我们有时就说,微分算子把一般的映射变成了矩阵计算问题.准确些说,则是一般的映射通过线性化以后归结为其切映射,而再通过引入坐标化成了很容易处理的矩阵计算.这正是线性化的真正的有力之处.

**注 1** 上面例如在引理 5 中我们说到 $X$ 是 $U$ 与 $U'$ 之间的微分同胚之类的话.怎样来判断 $X$ 是否微分同胚呢?由引理 5,$\mathrm{d}X\circ\mathrm{d}X^{-1}=I$,于是沿用上面的记号,有

$$\frac{\partial(x_1',\cdots,x_n')}{\partial(x_1,\cdots,x_n)}=\left[\frac{\partial(x_1,\cdots,x_n)}{\partial(x_1',\cdots,x_n')}\right]^{-1},$$

从而 $\dfrac{\partial(x_1,\cdots,x_n)}{\partial(x_1,\cdots,x_n)}$ 与 $\dfrac{\partial(x_1,\cdots,x_n)}{\partial(x_1',\cdots,x_n')}$ 均为非奇异的,即其行列式均不为 0.于是许多人可能设想,这也是 $X$ 成为微分同胚的充分条件.很遗憾,恰好不是,这只是 $X$ 成为局部微分同胚的充分必要条件.这一点在§5 中讲了反函数定理以后就清楚了.局部微分同胚和整体微分同胚是很不相同的.

**注 2** 上文中我们说到雅可比矩阵才是导数概念的合适的推广.从上一段讨论中可以看得更清楚:定理 4 讲的 $\mathrm{d}\Phi=\mathrm{d}g\circ\mathrm{d}f$,上面说了,正是雅可比矩阵的乘法公式,而它也就是复合函数的导数公式 $\dfrac{\mathrm{d}g[f(x)]}{\mathrm{d}x}=\dfrac{\mathrm{d}g}{\mathrm{d}y}\Big|_{y=f(x)}\cdot\dfrac{\mathrm{d}f}{\mathrm{d}x}$ 的推广.引理 5 的 $\mathrm{d}X^{-1}=(\mathrm{d}X)^{-1}$ 正是反函数导数公式 $\dfrac{\mathrm{d}y}{\mathrm{d}x}=1\Big/\dfrac{\mathrm{d}x}{\mathrm{d}y}$ 的推广.正如通常的微积分教本中在讲到反函数求导公式时,要加上 $\dfrac{\mathrm{d}y}{\mathrm{d}x}\neq0$, $\dfrac{\mathrm{d}x}{\mathrm{d}y}\neq0$ 这样的限制一样,引理 5 中要求 $x$ 与 $x'$ 均为彼此的可微函数,就自然导致了 $\partial$

向无穷大.

**注 3**　通常的微积分教本在讲到复合函数的微分时都会讲到一阶微分的形式不变性，其意思是，以一元函数为例.如果有复合函数 $\Phi(x)=g[f(x)]=(g\circ f)(x)$. 令 $y=g(x)$，则一方面有

$$\mathrm{d}\Phi=\Phi'(y)y'(x)\mathrm{d}x=\frac{\mathrm{d}\Phi}{\mathrm{d}x}\mathrm{d}x,$$

另一方面由 $y'(x)\mathrm{d}x=\mathrm{d}y$ 又有

$$\mathrm{d}\Phi=\Phi'(y)\mathrm{d}y.$$

所以不论 $y$ 是中间变量（如后一式）或 $x$ 是自变量（如前一式），$\mathrm{d}\Phi$ 的“表现形式”都是一样的.我们在这里不采用这个说法.因为 $\mathbf{R}^m$ 中可以有无穷多个坐标.用 $x$ 还是用 $x'$ 为坐标应该都是一样的，无所谓自变量与中间变量之别.对坐标 $x$ 有前一式，对坐标 $y$ 有后一式，其形状都是一样的，说明 $\mathrm{d}\Phi$ 其实与坐标的选取无关.其间的联系就是定理 4.所以真正重要的是定理 4，它表明 $\mathrm{d}\Phi$ 是一个内蕴的对象.而问题出现在，讨论高阶微分时就不再有这样有利的情况，而这正表明我们将揭开整个数学的新篇章了.

**4. 曲线、方向场、方向导数**　关于一阶微分与坐标选取的关系的讨论至此为止.下面我们将要讨论一些新的概念.在这以前，我们先从一个具体问题看一下，限制使用一定的坐标会带来什么样的问题.

设有曲线如图 3－2－3，a)，从直观看来，它在 $P$ 点有切线 $PQ$ 平行于 $y$ 轴，因而斜率为 $\infty$.如果按可微性的定义，它在 $P$ 点是不可微的.因为 $f(x+h)-f(x)=QP$，而 $h=QR$，因而当 $h\to 0$ 时，$QP/QR=+\infty$，所以找不到一个数 $A$ 使

$$f(x+h)-f(x)=Ah+o(h).$$

但是从几何上看，它完全是一个“很好”的曲线.如果我们把坐标轴逆时针倾斜一个小角度，或者说把曲线顺时针倾斜一个小角度如图 3－2－3，b)，则这个曲线立刻就成了一个可微函数的图像.因此，研究一条曲线的性质时，必须设法把曲线本身的性质与坐标的性质区别开来.

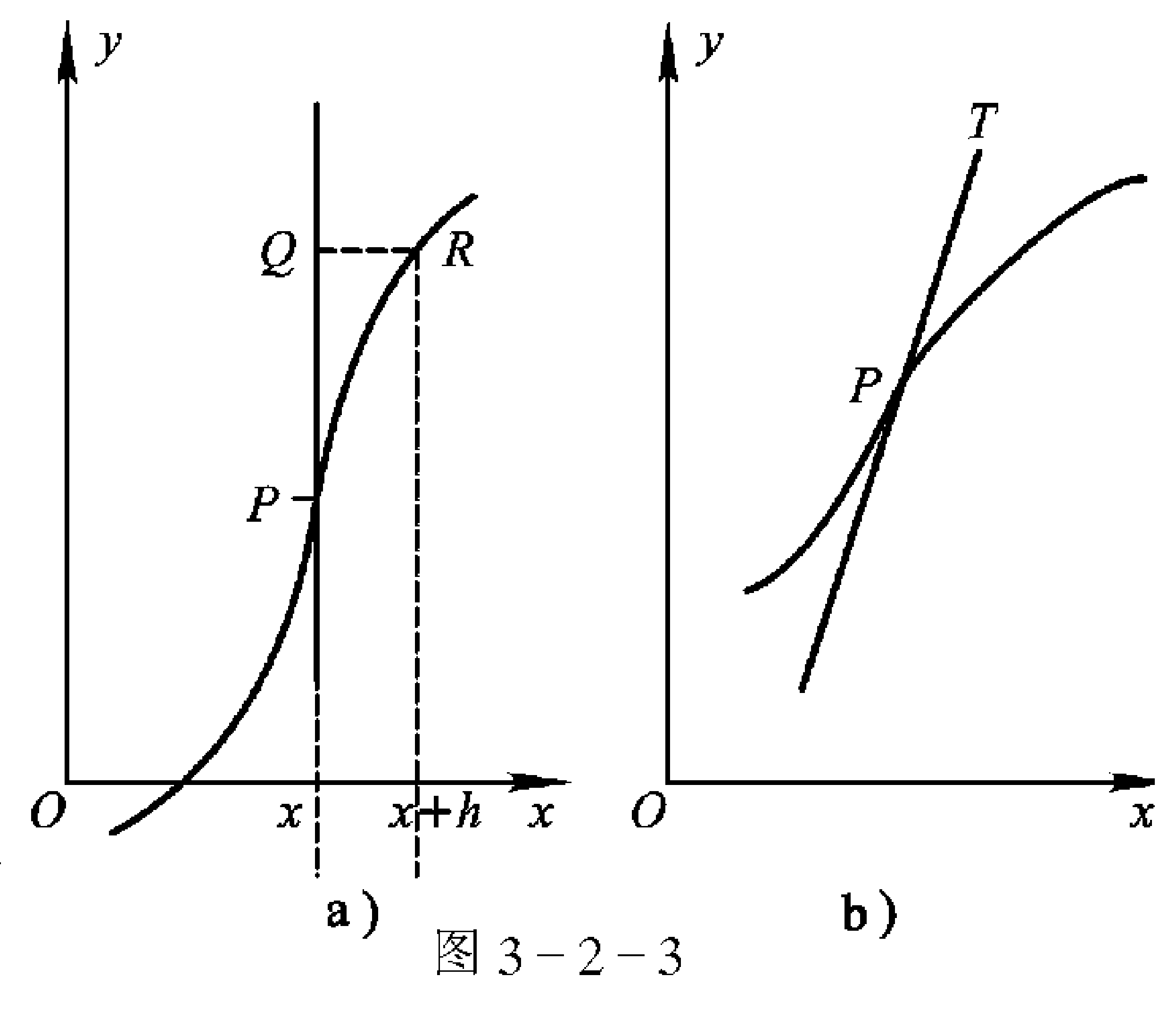


图 3－2－3

讲到这里，我们附带要提起一件在数学史上很有影响的事.19 世纪，当傅里叶讨论热的传导时，他得到了下面的级数.

$$\cos x-\frac{1}{3}\cos 3x+\frac{1}{5}\cos 5x-\cdots.$$

它是处处收敛的，其和例如在 $[-\pi,\pi]$ 之间是

$$s(x)=\begin{cases}\dfrac{\pi}{4}, & |x|<\dfrac{\pi}{2},\\ 0, & |x|=\dfrac{\pi}{2},\\ -\dfrac{\pi}{4}, & \dfrac{\pi}{2}<|x|<\pi.\end{cases}$$

(见第二章§4,图2-4-1).这是一个方波,从我们今天的观点看来是一个不连续函数.但是傅里叶不这么看,他认为极限函数的图像是若干条水平线段,即该图上的黑线,加上一些垂直线段,即该图上的虚线合成的,因而肯定是连续曲线.凡是可以笔不离纸地一笔画出来的图像,在傅里叶看来都是连续函数的图像(傅里叶发表他的结果早于柯西著名的《分析教程》——其中清楚地给出了连续性的定义——一年),再说,只要把图形逆时针稍微倾斜一点就会得到在 $x=\frac{\pi}{2}$ 处连续的图像——包括虚线——而顺时针倾斜一点又会得到在 $x=-\frac{\pi}{2}$ 处连续的图像.所以,有什么理由否定方波曲线是连续函数的图像呢?这个例子以及其他类似问题在19世纪初年引起了轩然大波,我们将在第五章中仔细讨论.也正是由此,狄利克雷才给出了函数的定义:有一个 $x$ 必有一个(注意,只有一个没有多个) $y$…….而对于方波,例如当 $x=\frac{\pi}{2}$ 时,虚线上的哪一点才算相应的函数值呢?所以它不是连续函数的图像.甚至任一条平行于 $y$ 轴的直线,例如 $x=1$ 都不可能是某函数(更不说连续函数了)的图像.终究我们回答不出来,当 $x=1$ 时有没有某个 $y$ 与之对应!狄利克雷的函数定义的一个现代的表述如下:一个函数 $f$ 即 $X\times Y=\{(x,y);x\in X,y\in Y\}$ 的一个子集 $F$,而且对同一个 $x$,不存在 $y_1\neq y_2$ 使 $(x,y_1),(x,y_2)\in F$.这里 $x$ 与 $y$ 明显地不对称.所以有许多书上强调说 $(x,y)$ 是一个有序偶:它不允许 $(x,y_1),(x,y_2)$ 同属于 $F$(当然,$y_1\neq y_2$)但允许 $(x_1,y),(x_2,y)$ 同属于 $F$,尽管 $x_1\neq x_2$. $F$ 中的 $(x,y)$ 之第一个分量 $x$ 之集 $\Omega$ 称为 $f$ 之定义域,第二个分量 $y$ 之集称为 $f$ 的值域. $F$ 中的 $(x,y)$ 之第二个分量 $y$ 全由第一个分量 $x$ 决定:有了一个 $x\in\Omega$,必有一个且只有一个 $y$ 使 $(x,y)\in F$,所以记作 $y=f(x)$. $F$ 也称为 $f$ 之图像.既然 $F$ 中 $x$ 与 $y$ 不对称,则不是任意曲线可以是某一函数的图像.上面的 $x=1$ 就是一个很好的例子.这样一来,对什么是曲线就需要定义.

**定义4**　$\mathbf{R}^n$ 中的 $C^k$ $(k\geqslant 1)$ 曲线就是一个由区间 $I=(a,b)\subset\mathbf{R}$ 到 $\mathbf{R}^n$ 中的 $C^k$ 映射 $\alpha$:

$$\alpha: I\to\mathbf{R}^n,\ t\ \mapsto\alpha(t)\in\mathbf{R}^n. \tag{27}$$

我们习惯是说这个映射的像为一曲线,但现在则说映射本身是曲线.映射的像时常称为曲线的迹(trace). $t$ 称为曲线的参数.如果令 $t=t(\tau)$,这里 $t(\tau)\in C^k$, $\tau\in(\alpha,\beta)$,而且 $t'(\tau)\neq 0$,曲线也可以写成

$$\alpha\circ t: J\to\mathbf{R}^n,\tau\ \mapsto\alpha(t(\tau))\in\mathbf{R}^n,$$

这里 $J=(\alpha,\beta)$(如果 $t'(\tau)<0$,则 $J$ 应为 $(\beta,\alpha)$).这称为曲线的重新参数化.有时,我们选曲线的弧长 $s$ 为参数,并且以 $t=a$ 作为弧长的起点.于是

$$s(t)=\int_a^t\sqrt{\sum_{i=1}^n x_i'^2(t)}\,\mathrm{d}t.$$

这样做时常会得到较简洁的结果.

如果给出一个 $C^k$ 曲线 $\alpha$,则 $\alpha'(t)$ 对每个 $t$ 都是 $\mathbf{R}^n$ 中的一个元,因此是一个向量.我们称它为此曲线在 $t$ 点的切向量.因为时常是以时间作为参数 $t$ 的,所以切向量也称为速度向量.而当我们确实在研究一个物理或力学问题,而 $t$ 又确实代表时间,则 $\alpha'(t)$ 时常写为 $\dot{\alpha}(t)$.这是牛顿留下的传统,沿用至今.

一个特别值得注意的事是，若在某点 $t_0$ 处 $\alpha'(t_0)=0$. 这时，切向量失去了方向！可是恰好是在这种点上，出现了种种内容极丰富的概念和方法，这种点现在都称为“奇点”. 一看到这个名词，读者会以为出了坏事：什么东西不连续了，变成无穷大了之类，完全是含有贬义. 不对！我们的定义中即已规定了 $\alpha(t)$ 的各个分量 $x_i(t), i=1,\cdots,n$，都属于 $C^k$，所谓“奇”就只是切向量失去了方向，因而产生了极为丰富的后果. 说句笑话，不妨把“奇点”理解为“出现奇迹的点”. 可惜，在这本书里我们完全不可能涉及这些问题了.

如果不涉及奇点，我们想讨论一下什么是两条曲线 $\alpha(t)$ 与 $\beta(t)$ 在 $t=t_0$ 处相切. 首先，它们在 $t=t_0$ 时相交：$\alpha(t_0)=\beta(t_0)$（记这个公共点为 $P$），由此我们知道 $\lim\limits_{t\to t_0}\|\alpha(t)-\beta(t)\|=0$. 现在我们要把 $\|\alpha(t)-\beta(t)\|$ 与 $|t-t_0|^l$（$l$ 为正整数）两个无穷小量比较. 如果 $\|\alpha(t)-\beta(t)\|=o(|t-t_0|^k)$（但不能是 $o(|t-t_0|^{k+1})$）. 这就说，两条曲线在 $P$ 点有 $k$ 阶接触（contact of order $k$）. 从直观上很清楚，只要 $k\geqslant 1$，这两条曲线就在 $P$ 点相切，因此有公共的切向量，也就是有公共的方向. 但是，仅仅只有通常微积分教材上讲的具有公共方向因而相切，还不能保证有 $k$ 阶接触. 一个明显的例子是 $\alpha(t)$ 与 $\lambda\alpha(t)$，这里 $\alpha(0)=0$，$\lambda$ 为实数，$\lambda\neq 1$，$\alpha(t)$ 与 $\lambda\alpha(t)$ 在 $t=0$ 点并非 1 阶接触. 提出“接触”这个概念的好处在于，上面我们都只在 $\mathbf{R}^n$ 中讨论曲线，而 $\mathbf{R}^n$ 中有坐标系和坐标轴，可以借此定义方向. 但是我们还要讨论例如曲面上的切线，例如地球表面上有子午线和纬圈，但是没有坐标轴. 在曲面上定义平行性极为困难，那么怎样定义方向呢？有了接触概念后问题就很容易解决了. 曲线在 $P$ 点相切的关系是一个等价关系. 利用这个等价关系把过 $P$ 点的曲线分类，每一个等价类就是一个方向. 现在看来，讲这样的话似乎多余，但到了将来，当需要在没有坐标轴的情况下讨论方向时，就会看到这个做法的好处了. 能对之进行线性运算的“对象”就是向量，于是这些等价类将可证明成为向量，而且会形成一个 $n$ 维线性空间. 这样一来，我们前面讲的概念和结果都可以大为推广了. 这在物理学上极为重要. 本书最后一章 §2 将详细讨论这些等价类何以成为一个 $n$ 维线性空间.

以下我们用 $l$ 表示曲线 $\alpha: I\to\mathbf{R}^n$ 在上述接触关系下的等价类. 上面已经说了每一个这样的等价类就定义了一个向量，它既有大小，又有方向（除了零向量之外）. 于是就可以把古典的微分学中的方向导数概念移植过来：

**定义 5**　设 $\alpha$ 是 $\mathbf{R}^m$ 中的一条 $C^k$ 曲线，$f:\mathbf{R}^m\to\mathbf{R}^n$ 是可微映射，我们称复合映射 $f\circ\alpha$ 的微分为 $f$ 沿 $\alpha$ 的方向导数：

$$\mathrm{d}(f\circ\alpha)=\mathrm{d}f\circ\mathrm{d}\alpha. \tag{28}$$

现在用坐标形式把它表示出来，设 $\alpha(t)={}^t(x_1(t),\cdots,x_m(t))$，$f={}^t(f_1,\cdots,f_n)$，而 $f_i=f_i(x_1,\cdots,x_m)$. 于是

$$\mathrm{d}\alpha={}^t(\dot{x}_1(t),\cdots,\dot{x}_m(t)).$$

$$\mathrm{d}f=\begin{pmatrix}\dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_m}\\ \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_m}\end{pmatrix}.$$

而

$$\mathrm{d}f\circ\mathrm{d}\alpha={}^{t}\left(\sum_{j=1}^{m}\frac{\partial f_1}{\partial x_j}\dot{x}_j(t),\cdots,\sum_{j=1}^{m}\frac{\partial f_n}{\partial x_j}\dot{x}_j(t)\right).\tag{29}$$

前面的上标 $t$ 表示转置,因此上式是一个竖向量.

这样看来,我们说可以证明 $\alpha$ 的等价类表示一个向量.现在就得找到一个将 $\alpha$ 与一个 $m$ 维向量对应起来的方式

$$\alpha\to(\dot{x}_1(t),\cdots,\dot{x}_m(t)).$$

它的方向固然是由上式决定的,它的大小也是由上式决定的. $\|\alpha\|=\left[\sum_{i=1}^{m}\dot{x}_i^2(t)\right]^{1/2}$. 现在问,如果用一个与 $\alpha$ 等价于 $t_0$ 点处的曲线 $\beta$ 代替 $\alpha$,结果如何?上面已说了,$\alpha$ 与 $\beta$ 在 $t_0$ 点等价,即二者在 $t_0$ 点有 $k$ 阶接触,$k\geqslant 1$. 因此一方面 $\alpha(t_0)=\beta(t_0)$,即通过同一点,另一方面 $\|\alpha(t)-\beta(t)\|=o(|t-t_0|^k)$,所以 $\dot{\alpha}(t_0)=\dot{\beta}(t_0)$,即都可以用 $(\dot{x}_1(t_0),\cdots,\dot{x}_m(t_0))$表示.由(29)即知 $\mathrm{d}f\circ\mathrm{d}\alpha=\mathrm{d}f\circ\mathrm{d}\beta(t=t_0)$.所以定义中的方向导数值其实只依赖于 $\alpha$ 的等价类 $l$. $l$ 作为一个向量,其方向就是$(\dot{x}_1(t),\cdots,\dot{x}_m(t))$的方向,其大小则是 $\|l\|=\left[\sum_{i=1}^{m}\dot{x}_i^2(t)\right]^{1/2}$. 以下我们用$\frac{\partial f}{\partial l}(P)$来表示在 $P$ 点(即 $t=t_0$ 处)沿 $\alpha$ 之方向导数,而且就说成是"沿 $l$ 方向的"方向导数.

这里的讲法和通常微积分教本中的讲法有些不同.在那里,讲到方向导数时,是用一个单位向量 $u=(u_1,\cdots,u_m)$来表示此方向.然后设 $P$ 点坐标是$(x_1,\cdots,x_m)$,作通过 $P$ 点的以 $u$ 为方向余弦的直线 $\alpha$: $t\to(x_1+tu_1,\cdots,x_m+tu_m)$,然后在 $t=0$ 处对 $f$ 求$\frac{\partial f}{\partial\alpha}$.利用(29)就有

$$\frac{\partial f}{\partial\alpha}=\lim_{t\to 0}\frac{f(x+tu)-f(x)}{t}.$$

为什么我们不采用通常的讲法?一是因为可能需要作坐标变换,而原来的直线可能变成曲线(如果坐标系总画成直角坐标系的话).而甚至可能根本没有直线,例如在曲面上.而通过重新参数化 $t=t(\tau)$以后 $\|u\|=1$ 的这一条件也可能不成立.这一点其实前面也讲过.如果把参数 $t$ 看成时间,则切向量 $\dot{x}(t)$应理解为速度.这样一来,$\frac{\partial f}{\partial\alpha}$就成了:当点 $x$ 以速度 $\dot{x}$运动时,$f$ 对时间的变化率 $\mathrm{d}f\circ\mathrm{d}\alpha$ 就是方向导数. $\mathrm{d}f\circ\mathrm{d}\alpha$ 就理解为:首先 $t$ 的变化造成了 $x$ 位置之变化 $\mathrm{d}\alpha$(其变率为$(\dot{x}_1(t),\cdots,\dot{x}_m(t))$),位置的变化再造成 $f$ 的变化 $\mathrm{d}f$ $\left(\text{其变化率为}\left(\frac{\partial f}{\partial x_1},\cdots,\frac{\partial f}{\partial x_m}\right)\right)$.二者合成即得 $\mathrm{d}f\circ\mathrm{d}\alpha$.

进一步可以定义向量场.设 $\Omega\subset\mathbf{R}^m$ 为一开集,而在 $\Omega$ 内之每一点 $x$ 处都定义一个向量 $X(x)=\sum_{i=1}^{m}X_i(x)\frac{\partial}{\partial x^i}$,我们就说有一个向量场.如果 $X_i(x)\in C^k(\Omega)$,就说 $X(x)$是一个 $C^k$ 向量场.常微分方程理论告诉我们,这时经过 $\Omega$ 中任一点 $P$ 均可找到唯一一条曲线 $\alpha(t)$: $I\to\mathbf{R}^m$,使 $\alpha(t_0)=x_0$ 即为 $P$ 点之坐标,而沿此曲线恒有 $\dot{\alpha}(t)=X(\alpha(t))$.这条曲线就称为此向量场经过 $P$ 点的积分曲线.以上的讨论中我们要利用 $k\geqslant 1$ 这一事实,才能保证经过任一 $P$ 点均有唯一的积分曲线通过.

既然有了向量场，也就有了一个方向（在现在的问题中，即是 $(X_1,\cdots,X_m)$ 的方向），也就是在此点的 $\dot{\alpha}(t)$，所以也就可以定义这个方向的方向导数和(28)式．我们约定，把这个方向导数算子记作 $\mathrm{D}_{X(x)}$．即是说按下式定义它：

$$\mathrm{D}_{X(x)}f(x)=\mathrm{d}f(x)\circ\mathrm{d}\alpha(x). \tag{30}$$

这里 $\mathrm{d}\alpha(x)$ 其实应该写成 $\mathrm{d}\alpha(t)$，$t$ 是相应于 $x$ 的参数值．$\mathrm{D}_{X(x)}$ 是一个很重要的算子，其特点如下．

上面我们已经说过，可以把接触于某点的曲线所形成的等价类当作一个向量．如果在这个等价类中选一个代表元，即曲线 $\alpha(t)$，则这个向量有一个坐标表示 $\dot{\alpha}(t)=(\dot{\alpha}_1(t),\cdots,\dot{\alpha}_m(t))$，$t$ 是该点 $P$ 的相应的参数值．按同样的方法，方向导数算子 $\mathrm{D}_{X(x)}$ 也是一个向量．因为由(30)式，$\mathrm{D}_{X(x)}$ 作用于一个 $n$ 维向量 $f(x)\in\mathbf{R}^n$ 后得出

$$\begin{aligned}\mathrm{D}_{X(x)}f(x)&=\mathrm{d}f(x)\circ\mathrm{d}\alpha(x)\\&={}^t\left(\sum_{j=1}^m X_j(x)\frac{\partial f_1}{\partial x_j},\cdots,\sum_{j=1}^m X_j\frac{\partial f_m}{\partial x_j}\right).\end{aligned} \tag{31}$$

所以

$$\mathrm{D}_{X(x)}\sim X(x)=\sum_{j=1}^m X_j(x)\frac{\partial}{\partial x^j}. \tag{32}$$

正如我们可以认为曲线的等价类就是向量一样，也可以认为微分算子 $X(x)=\sum_{j=1}^m X_j(x)\frac{\partial}{\partial x^i}$ 是一个向量．反过来也对，正因为这样，$\mathrm{D}_{X(x)}$ 也可以写成 $\mathrm{D}_{\frac{\partial}{\partial x_1}}$ 等等．但是认真想一下，这样做的“毛病”无非是使我们感到怪怪的．如果说 $\mathrm{D}_1=\mathrm{D}_{x_1}$ 表示沿 $x_1$ 方向的方向导数，就是令自变量在 $x_1$ 方向得到一个改变量 $\Delta x_1$，然后计算函数 $f$ 的相应的改变量 $\Delta f$，并令 $\mathrm{D}_{x_1}=\lim\limits_{\Delta x_1\to 0}\frac{\Delta f}{\Delta x_1}$，这样的说法谁也不会奇怪．可是现在我们有了 $\mathrm{D}_{\frac{\partial}{\partial x_1}}$，D 的下标是 $\frac{\partial}{\partial x_1}$，我们怎么能做出它的改变量 $\Delta\left(\frac{\partial}{\partial x_1}\right)$ 呢？但是再回过来看看(32)式，这些疑问会逐步释然于怀了．我们说，把微分算子 $X(x)=\sum_{j=1}^m X_j(x)\frac{\partial}{\partial x_j}$ 就看成向量 $(X_1(x),\cdots,X_m(x))$，这一点大概读者会能接受得了，但是既然承认了这一点，也就应该承认 $\frac{\partial}{\partial x_1},\cdots,\frac{\partial}{\partial x_m}$ 是基底，因此例如 $\frac{\partial}{\partial x_1}$ 也是一个向量，即在 $x_1$ 方向上分量为 1，其他方向上分量为 0 的向量．所以它的坐标表示就是 $\frac{\partial}{\partial x_1}=\sum_{j=1}^m X_j(x)\frac{\partial}{\partial x_j}$，$X_1=1$，$X_2=\cdots=X_m=0$．我们也就可以写作 $\frac{\partial}{\partial x_1}=(1,0,\cdots,0)$．这样一来，$\mathrm{D}_{\frac{\partial}{\partial x_1}}$ 就表示对于向量 $(1,0,0)$ 求方向导数，也就是 $\mathrm{D}_1$！既然 $\mathrm{D}_{\frac{\partial}{\partial x_1}}=\mathrm{D}_1$，又何必兴师动众用那么复杂的记号呢？

$\mathrm{D}_{\sum\limits_j X_j(x)\frac{\partial}{\partial x_j}}=\mathrm{D}_{X(x)}$ 这样的记号有许多重要性质：

(1) 首先它要作用于函数上．函数是“被作用者”．这时它有两个性质．一是线性：

$$\mathrm{D}_X(C_1f_1+C_2f_2)=C_1\mathrm{D}_Xf_1+C_2\mathrm{D}_Xf_2, \tag{33}$$

$C_1, C_2$ 是常数. 二是莱布尼茨性质，即

$$\mathrm{D}_X(f_1 \cdot f_2) = f_1 \mathrm{D}_X f_2 + f_2 \mathrm{D}_X f_1. \tag{34}$$

这一点由(30)式直接验证即知.

(2) 其次，它是按 $X(x)$ 的要求作用于函数的. $X(x)$ 可是说是“作用者”. 而这些作用者又有线性结构，那么 $\mathrm{D}_{X(x)}$ 对这种结构又有什么性质呢？这是十分值得注意的. 有

**定理 6** 设 $g_1(x), g_2(x)$ 是 $\Omega$ 中的 $C^k$ 函数，则对 $\Omega$ 中的 $C^k$ 向量场 $X_1(x), X_2(x)$ 恒有

$$\mathrm{D}_{g_1X_1+g_2X_2} f(x) = g_1 \mathrm{D}_{X_1}(f)(x) + g_2 \mathrm{D}_{X_2}(f)(x). \tag{35}$$

$f$ 是 $\Omega$ 上的可微函数.

**证** 因为 $\mathrm{D}_X f(x) = \mathrm{d}f \circ \dot{\alpha}(x)$. 这里 $\dot{\alpha}(x)$ 即是方向场 $X$ 在 $x$ 之值. 所以令 $X_1, X_2$ 相应的积分曲线是 $\alpha_1, \alpha_2$，则 $g_1X_1 + g_2X_2$ 在 $x$ 之值应为 $g_1(x)\dot{\alpha}_1(x) + g_2(x)\dot{\alpha}_2(x)$：

$$\begin{aligned} D_{g_1X_1+g_2X_2}(f)(x) &= g_1(x)\mathrm{d}f \circ \dot{\alpha}_1(x) + g_2(x)\mathrm{d}f \circ \dot{\alpha}_2(x) \\ &= g_1(x)\mathrm{D}_{X_1}(f)(x) + g_2(x)\mathrm{D}_{X_2}(f)(x). \end{aligned}$$

(35)式得证. 这里我们应用了以下事实：$\mathrm{d}f$ 实际上就以乘以 $n \times m$ 矩阵 $\dfrac{\partial(f_1, \cdots, f_n)}{\partial(x_1, \cdots, x_m)}$. $\dot{\alpha}_1(x)$ 也是一个 $m \times 1$ 矩阵，$g_1(x)$ 则只是一个标量因子，所以

$$\frac{\partial(f_1, \cdots, f_n)}{\partial(x_1, \cdots, x_m)} g_1(x)\dot{\alpha}_1(x) = g_1(x)\frac{\partial(f_1, \cdots, f_n)}{\partial(x_1, \cdots, x_m)}\dot{\alpha}_1(x).$$

总结起来，$\mathrm{D}_X f$ 中其实含有两个“成分”，一是“被作用者” $f$. $\mathrm{D}_X f$ 对于它既是线性的，又适合莱布尼茨法则；二是“作用者” $X$，$\mathrm{D}_X f$ 按定理 6 其实还不只是线性地依赖于它. 因为(35)式只要当 $g_1, g_2$ 取常数时成立，就已经表明 $\mathrm{D}_X f$“线性”依赖于 $X$ 了，可是现在它甚至对 $C^k$ 函数 $g_1(x), g_2(x)$ 也成立. 可见这个依赖性还超过“线性”. 这是一种新的代数结构，也是有用的.

可是，我们费了这许多口舌都是为了另一件事. 现在我们只是对一个函数 $f$(映射 $f: \mathbf{R}^m \to \mathbf{R}^n$ 则是 $n$ 个函数)作方向导数. 而更严重的问题却是要对更复杂的对象，例如对一个向量场 $Y$ 可作方向导数 $\mathrm{D}_X Y$. 而且不是在 $\mathbf{R}^m$ 的某一个区域 $\Omega$ 上作，甚至是在一个曲面以至更一般的弯曲的空间上作，而且这在几何和物理学上都十分重要. 有鉴于此，尽管本书讲不到这些问题，预为读者计，现在有一个“心理准备”，一旦遇到这类问题会有似曾相识之感. 这是会有好处的，值得现在花点力气.

**5. 高阶微分** 本节最后，我们简单讨论一下一个映射 $f: \Omega \subset \mathbf{R}^m \to \mathbf{R}^n$ 的高阶微分. 作为一个启发我们先看最简单的一元函数 $f(x)$ 的二阶微分. 现在 $\Omega$ 将成为 $\mathbf{R}$ 的一个区间，$\mathbf{R}^n$ 也成为 $\mathbf{R}$. 首先看一阶微分 $\mathrm{d}f$，在 $x \in \Omega$ 点，一阶微分 $\mathrm{d}f(x)$ 就是 $f(x+h) - f(x) = \Delta_h f(x)$ 关于 $h$ 的线性部分：$\Delta_h f(x) = \mathrm{d}f(x) \cdot h + o(h)$. 于是 $\mathrm{d}f(x)$ 又是一个由 $x \in \Omega$ 到 $\mathrm{d}f(x)$ 的映射. 不过这里 $\mathrm{d}f(x)$ 是一个线性映射，它把 $\Omega$ 的切空间，即仍为 $\mathbf{R}$(前面说过这就是 $h$ 的空间)映为靶空间 $\mathbf{R}^n$ 的切空间(仍为 $\mathbf{R}^n$，现在 $n=1$). 也就是 $\mathbf{R}^1 \to \mathbf{R}^n$ 的线性映射(即一阶矩阵). $n=1$ 时，一阶矩阵就是一个数，这个数即是古典意义下的导数 $f'(x)$. 因为这个数依赖于 $x$，所以是一个函数. 但是这个函数不是线性函数. 现在再求这个函数的微分，就得到二阶微分. 因此，我们又要讨论

$f'(x+h)-f'(x)$. 但是有什么理由要求现在的自变量增量 $\Delta x$ 仍旧是 $h$ 呢？因此，现在可以设 $\Delta x=k$ 而 $h$ 和 $k$ 是独立的. 由这个思想出发，我们应该这样来讨论 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$ 的二阶微分. 首先假设有一个微分 $\mathrm{d}f(x)$（有的书上则把微分写成 $T_xf$ 或 $(Tf)(x)$，大概是因为 $\mathrm{d}f$ 作为一个映射称为 $f$ 的切映射，而 $T$ 正是“切”(tangent)的第一个字母)，即有

$$\Delta_h f(x)=f(x+h)-f(x)=(\mathrm{d}f)(x)\cdot h+o(h).$$

$(\mathrm{d}f)(x)$ 是由 $\Omega$ 到 $n\times m$ 实矩阵的空间的映射，即是：有一点 $x\in\Omega$，即有一个这样的矩阵，所以是矩阵值函数，它当然一般是非线性的. $n\times m$ 实矩阵有 $nm$ 个实数元，反过来，凡有 $nm$ 个实数，总可以写成一个 $n\times m$ 实矩阵，所以这类矩阵所成的空间是 $\mathbf{R}^{nm}$. 但是这样来看矩阵空间并不好，因为还得规定一个规矩：哪一个实数放在第 $k$ 行 $l$ 列（但以后我们在讨论正交矩阵空间时确是这样做的，即已隐含地规定好了这样一个规矩），所以我们把每个这样的矩阵都看成由 $\mathbf{R}^m$ 到 $\mathbf{R}^n$ 的线性映射，并把这种映射空间记作 $L(\mathbf{R}^m,\mathbf{R}^n)$. 如果要排除对记号 $L(\mathbf{R}^m,\mathbf{R}^n)$ 的生疏感，就把它看成矩阵空间，甚至是赋有某种次序规定的 $\mathbf{R}^{nm}$. 总之，它是一个 $nm$ 维线性空间. $(\mathrm{d}f)(x):\Omega\to L(\mathbf{R}^m,\mathbf{R}^n)$. 于是我们可以考虑 $(\mathrm{d}f)(x)$ 的微分 $\mathrm{d}(\mathrm{d}f)(x)$. 即由 $\Omega$ 之切空间 $\mathbf{R}^m$ 到 $L(\mathbf{R}^m,\mathbf{R}^n)$ 的切空间的线性映射：

$$(\mathrm{d}f)(x+k)-(\mathrm{d}f)(x)=\mathrm{d}(\mathrm{d}f)\circ k+o(k).$$

但因 $L(\mathbf{R}^m,\mathbf{R}^n)$ 是一个线性空间，它的切空间就是其自身. 所以 $\mathrm{d}(\mathrm{d}f)(x)$ 对每一点 $x\in\Omega$ 定义了由 $\mathbf{R}^m$ 到 $L(\mathbf{R}^m,\mathbf{R}^n)$ 的线性映射：$\mathrm{d}(\mathrm{d}f)(x):\Omega\to L(\mathbf{R}^m,\ L(\mathbf{R}^m,\mathbf{R}^n))$. 所以我们有

**定义 6** 若 $(\mathrm{d}f)(x):\mathbf{R}^m\to L(\mathbf{R}^m,\mathbf{R}^n)$ 为连续可微，则其微分是 $\Omega\to L(\mathbf{R}^m,\ L(\mathbf{R}^m,\mathbf{R}^n))$ 的线性映射，称为 $f$ 在 $\Omega$ 中的二阶微分，记作 $\mathrm{d}^2f(x)$. 这种映射称为属于 $C^2(\Omega)$ 类.

我们要证明，$L(\mathbf{R}^m,L(\mathbf{R}^m,\mathbf{R}^n))$ 即双线性映射 $L(\mathbf{R}^m,\mathbf{R}^m;\mathbf{R}^n)$ 的空间. 所谓双线性映射即映射 $\mathbf{R}^m\times\mathbf{R}^m\to\mathbf{R}^n$，$(h,k)\mapsto l(h,k)$，而对每个分量 $h$ 与 $k$ 分别为线性的映射：

$$l(h,c_1k_1+c_2k_2)=c_1l(h,k_1)+c_2l(h,k_2),$$
$$l(c_1h_1+c_2h_2,k)=c_1l(h_1,k)+c_2l(h_2,k).$$

这里 $c_1,c_2$ 是常数. 于是有

**引理 7** 我们有同构关系：$L(\mathbf{R}^m,L(\mathbf{R}^m,\mathbf{R}^n))\simeq L(\mathbf{R}^m,\mathbf{R}^m;\mathbf{R}^n)$.

**证** 从左方取一个元 $l$. 对 $h\in\mathbf{R}^m$，$l\cdot h\in L(\mathbf{R}^m,\mathbf{R}^n)$，$l\cdot h$ 对于 $h$ 是线性的. 再从 $\mathbf{R}^m$ 中取一个 $k$，则 $(l\cdot h)(k)\in\mathbf{R}^n$ 令 $(l\cdot h)\cdot(k)=g$. 此式对于 $k$ 也是线性的. 显然，我们是从一对元 $(h,k)\in\mathbf{R}^m\times\mathbf{R}^m$ 得出了元 $g\in\mathbf{R}^n$. 而且很容易看出，这个对应是双线性的. 所以我们得到右方一个元 $\tilde{l}$，而 $(l\cdot h)\cdot(k)=\tilde{l}(h,k)$. 这里 $h$ 在左右双方都取自靠左的一个 $\mathbf{R}^m$. 反之，设有 $\tilde{l}\in L(\mathbf{R}^m,\mathbf{R}^m;\mathbf{R}^n)$. 从右方靠左的 $\mathbf{R}^m$ 中取一个元 $h$，则将得到 $\tilde{l}(h\cdot)$，需再从靠右的 $\mathbf{R}^m$ 中取元 $k$，代入 $\tilde{l}(h,\cdot)$ 中才能得到 $\mathbf{R}^n$ 中一个元 $g$：$\tilde{l}(h,k)=g$. 可见 $\tilde{l}(h,\cdot)$ 是映 $k\in\mathbf{R}^m$ 为 $g\in\mathbf{R}^n$ 的线性映射，这个映射的线性容易证明. 记它为 $(l\cdot h)(\cdot)$，则 $l\cdot h\in L(\mathbf{R}^m,\mathbf{R}^n)$，而 $\tilde{l}(h,g)=(l\cdot h)(g)$. 总之我们在 $L(\mathbf{R}^m,L(\mathbf{R}^m,\mathbf{R}^n))$ 与 $L(\mathbf{R}^m,\mathbf{R}^m;\mathbf{R}^n)$ 之间建立了一个一一对应关系，而且易见这个一一对应保持线性结构不变，因此得到引理 7 的证明.

现在利用坐标表示讨论映射 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$ 为 $C^2$ 映射之条件. 为方便起见，我们将分别讨论 $f$ 之每一个分量：$f=(f_1,\cdots,f_n)$. 于是设 $\mathbf{R}^m$ 与 $\mathbf{R}^n$ 中都已有了坐标 $x=(x_1,\cdots,x_m)$，$y=$

$(y_1, \cdots, y_n)$，而

$$y_i = f_i(x_1, \cdots, x_m) \qquad i = 1, 2, \cdots, n.$$

现在我们只讨论 $f_1$，并将它写成 $f$，取 $(h, k) \in \mathbf{R}^m$. 如果记一阶微分映射为 $\mathrm{d}f$，则

$$f(x+h) - f(x) = (\mathrm{d}f)(x) \cdot h + o(h).$$

但是，由定义 5，$(\mathrm{d}f(x) \cdot h)$ 正是 $f$ 沿方向 $h$ 的方向导数，用 (30) 中的记号把它写成 $(\mathrm{D}_h f)(x)$. 于是由定义 6 知，所谓 $f \in C^2$ 就是指对 $\Omega$ 中任意 $x$ 以及任意 $h \in \mathbf{R}^m$，方向导数 $(\mathrm{D}_h f)(x) \in C^1$. 再取 $k \in \mathbf{R}^m$，我们又应有

$(\mathrm{D}_h f)(x+k) - (\mathrm{D}_h f)(x) = \mathrm{d}(\mathrm{D}_h f)(x) k + o(k)$. 再用一次方向导数记号：$\mathrm{d}(\mathrm{D}_h f)(x) k = \mathrm{D}_k(\mathrm{D}_h(f))(x)$，而由 $C^2$ 之定义知此式应为 $\Omega$ 中的连续函数. 可见这就是 $f: \Omega \to \mathbf{R}$ 为 $C^2$ 映射的充分必要条件. 于是有

**定理 8** $f: \Omega \to \mathbf{R}$ 为 $C^2$ 映射的充分必要条件是在取定任意坐标 $x = (x_1, \cdots, x_m)$ 后，古典的二阶偏导数 $\dfrac{\partial^2 f}{\partial x_i \partial x_i}$ 为 $\Omega$ 中的连续函数.

**证** 在 $\mathbf{R}^m$ 中取定坐标 $x = (x_1, \cdots, x_m)$. 也是由定理 6 前面的讨论，$\dfrac{\partial}{\partial x_1}, \cdots, \dfrac{\partial}{\partial x_m}$ 成为 $\mathbf{R}^m$ 的一个基底，而任意 $h, k$ 分别可以写为 $h = \sum\limits_{i=1}^{m} h_i \dfrac{\partial}{\partial x_i}$，$k = \sum\limits_{j=1}^{m} k_j \dfrac{\partial}{\partial x_j}$. 而由定理 6 及 (32) 式

$$(\mathrm{d}^2 f)(x) h \cdot k = \mathrm{D}_k(\mathrm{D}_h f)(x) = \sum_{i,j=1}^{m} h_i k_j \left[\mathrm{D}_{\frac{\partial}{\partial x_j}} \left(\mathrm{D}_{\frac{\partial}{\partial x_i}} f\right)(x)\right].$$

于是 $f \in C^2$ 之充分必要条件就是：对任意 $h, k \in \mathbf{R}^m$，$(\mathrm{d}^2 f)(x) hk$ 是 $x \in \Omega$ 的连续函数. 但此式是 $h, k$ 的双线性形式，它对 $x \in \Omega$ 为古典的连续函数的充分必要条件就是其系数 $\mathrm{D}_{\frac{\partial}{\partial x_j}}(\mathrm{D}_{\frac{\partial}{\partial x_i}} f)(x)$ 是 $x$ 的连续函数. 但是在定理 5 前的讨论中已经看到 $\mathrm{D}_{\frac{\partial}{\partial x_j}}(\mathrm{D}_{\frac{\partial}{\partial x_i}})(x)$ 即古典的二阶偏导数 $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$. 故定理得证.

以上我们是对 $f: \Omega \to \mathbf{R}^1$ 的情况证明定理 8 的，但是由前面的说明可知对 $f: \Omega \to \mathbf{R}^n$ 也有类似的结论.

还有一个重要的说明. 对于 $f: \Omega \to \mathbf{R}^1$ 的情况，$\mathrm{d}^2 f(x)$ 就是一个双线性形式

$$(\mathrm{d}^2 f)(x) hk = \sum_{i,j=1}^{m} h_i k_j \frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{36}$$

一个双线性形式与定义它的矩阵是一一对应的，这个矩阵在我们的情况下称为黑塞矩阵 (Hessian, Hesse matrix).

$$\mathrm{Hess}_x f(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_m} \\ & & \\ \dfrac{\partial^2 f}{\partial x_m \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_m^2} \end{pmatrix}. \tag{37}$$

也有一些文献中称此矩阵的行列式为 Hessian.

这一段一开始我们就提出，没有理由让 $\Delta x$ 两次都取成相同的 $h$. 因此，我们第一次取 $\Delta x =$

$h$，第二次取 $\Delta x=k$，而得到了一个双线性形式(36)．如果取 $h=k$，则双线性形式(36)将变成一个二次型

$$\sum h_i h_j \frac{\partial^2}{\partial x_i \partial x_j}.$$

反过来，由二次型变成双线性形式亦非难事．线性代数里面讲过这一点．但是如果考虑更高阶的微分，事情就远非这么简单．这也是为什么我们坚持两个 $\Delta x$ 各不相同，而分别为 $h, k$ 的原因．

关于高阶导数在通常的微积分教材中都会介绍一个重要的定理，即是若 $f(x,y)\in C^2$，必有

$$\frac{\partial^2 f}{\partial x \partial y}=\frac{\partial^2 f}{\partial y \partial x}. \tag{38}$$

然后一般会举出一两个反例，说明若 $f(x,y)\in C^2$ 这个条件不成立，则上式也可能不成立．有时，人们就会去想，能否找一个较弱的条件保证上式成立．但忽视了这个等式是否有深远的后果．我们要指出，这个结果虽然证明十分简单，却是一个重要的结果，我们将在最后一章讨论它．若它不成立，将意味着什么也是很有意思的事，这一点本书完全不能涉及了．正因为它如此重要，我们将在本节中一直采用的与坐标无关的框架下去讨论它．

若(38)成立，则黑塞矩阵(37)成为一个对称矩阵，而双线性形式成为一个对称的双线性形式．本来，按线性代数的讲法，一个双线性形式 $\sum\limits_{i,j=1}^{n} a_{ij}h_i k_j$ 当 $a_{ij}=a_{ji}$ 恒成立时就称为对称的双线性形式．由它就可以产生二次型 $\sum\limits_{i,j=1}^{m} a_{ij}h_i k_j$．所以二次型总要适合上述对称性条件 $a_{ij}=a_{ji}$．而一般说来，双线性形式则不一定是对称的．我们即将证明(38)式．在未证明之前，(36)式虽是一个双线性形式，但不能直接在其中令 $h=k$ 而得二次型 $\sum\limits_{i,j=1}^{n} h_i h_j \dfrac{\partial^2 f}{\partial x_i \partial x_j}$．对这个问题以前我们没有很严格地讨论过，在此再提醒一下．

上面已说了，当 $f\in C^2$ 时，(38)成立，而(36)成为一个对称的双线性形式．现在我们要证明

**定理 9** 设 $f:\Omega\to\mathbf{R}^2$ 是 $C^2$ 映射，则 $(\mathrm{d}^2 f)(x)hk$ 是对称的，即 $(\mathrm{d}^2 f)(x)hk=(\mathrm{d}^2 f)(x)kh$．

**证** 这个定理虽然是在与坐标无关的框架下表述的，但它的内容和证法都与通用的微积分教本中证明(38)式是一样的，就是用差商趋向微商，而对差商则利用拉格朗日格公式．所以我们先引入差分

$$(\Delta_h f)(x)=f(x+h)-f(x).$$

为了求二阶微分，我们再引入差分 $\Delta_k$，$k$ 与 $h$ 无关．

$$\begin{aligned}\Delta_k(\Delta_h f)(x)&=(\Delta_h f)(x+k)-(\Delta_h f)(x)\\&=f(x+h+k)-f(x+k)-f(x+h)+f(x).\end{aligned} \tag{39}$$

很明显，(39)对于 $h$ 与 $k$ 是对称的，所以

$$\Delta_h(\Delta_k f)(x)=\Delta_k(\Delta_h f)(x).$$

由差分转向微分时，在通常的微分教本中是用拉格朗日公式，但我们在讲定理 2 前就指出，$n>1$ 时是没有拉格朗日公式那样的中值定理的，而我们要用一个不等式来代替它，这个不等式就是定理 2，也称为中值定理．其中的 $M$ 可以取为连接 $a, b$ 两点的线段上微分(作为一个映射)范数的上确界．于是在考虑 $f(x+h)-f(x)$ 时，我们可以更精确地来估计它．我们本来想估计

$f(x+h)-f(x)$. 现在则更精确一点来估计 $\Phi(t)=f[x+th]-(\mathrm{d}f)(x)th$，当 $t$ 在 $[0,1]$ 中变动时 $x+t(y-x)$ 就沿连接 $x$ 与 $x+h$ 的线段从 $x$ 变到 $x+h$，所以可以应用定理 2，这样做的好处在于，$\Phi(1)-\Phi(0)=f(x+h)-f(x)-(\mathrm{d}f)(x)h$，所以我们不仅可以估计到 $f(x+h)-f(x)$，而且估计到了 $f(x+h)-f(x)-(\mathrm{d}f)(x)h$. 注意到我们假设了 $f\in C^2$，所以 $(\mathrm{d}f)(x)\in C^1$，这样 $\Phi(t)\in C^1$. 所以应用定理 2 是合法的. $\Phi(t)$ 对于 $t$ 的微分是

$$f'(x+th)h-(\mathrm{d}f)(x)h,$$

所以定理 2 中的 $M$ 可以取为 $\sup\limits_{0\leqslant t\leqslant 1}\|f'(x+th)h-(\mathrm{d}f)(x)h\|$.

用完全同样的方法，在考虑二阶微分时，我们考虑

$$\Phi(t,s)=f(x+th+sk)-(\mathrm{d}^2f)(x)(th,sk),$$

$(\mathrm{d}^2f)(x)$ 是一个双线性形式，右方后一项就是它在 $(th,sk)$ 上之值，对 $t$ 和 $s$ 分别应用一次上面的程序，就会得到

$$\begin{aligned}\|\Delta_k\Delta_hf(x)-(\mathrm{d}^2f)(x)(h,k)\|&\leqslant\sup\|(\mathrm{d}^2f)(x+th+sh)(h,k)-(\mathrm{d}^2f)(x)(h,k)\|\\&=o(1)(\|h\|\cdot\|k\|)\end{aligned}$$

在上式中对换 $h$ 与 $k$，由于 $\Delta_k\Delta_hf(x)=\Delta_h\Delta_kf(x)$ 所以有

$$\begin{aligned}&\|(\mathrm{d}^2f)(x)(k,h)-(\mathrm{d}^2f)(x)(h,k)\|\\=&\|\Delta_k\Delta_hf(x)-(\mathrm{d}^2f)(x)(h,k)-\Delta_h\Delta_kf(x)+(\mathrm{d}^2f)(x)(k,h)\|\\=&o(1)(\|h\|\cdot\|k\|).\end{aligned}$$

但是式左是 $h,k$ 的双线性形式，它不可能是比 $\|h\|\cdot\|k\|$ 更高阶的无穷小量，除非恒等于 0. 因此对任意 $h,k\in\mathbf{R}^m$

$$(\mathrm{d}^2f)(x)(k,h)=(\mathrm{d}^2f)(x)(h,k).$$

定理证毕.

由此可见，拉格朗日公式是多么好的东西. 少了它带给我们多少麻烦！

以上我们讨论了二阶微分，对更高阶的微分可以类似地讨论. 我们不再重复，只把主要的结论归结如下：

1. 映射 $f:\Omega\to\mathbf{R}^n\ (\Omega\subset\mathbf{R}^m)$ 为 $C^k$ 类映射之定义是：$f\in C^k$，而且 $\mathrm{d}f\in C^{k-1}$.

2. 上述 $f$ 的 $k$ 阶微分 $\mathrm{d}^kf$ 是由 $\Omega$ 到 $k$ 阶多重线性映射 $\Omega\to L(\underbrace{\mathbf{R}^m,\mathbf{R}^m,\cdots,\mathbf{R}^m}_{k\text{ 个因子}};\mathbf{R}^n))$ 空间的映射. 即是说，对每一点 $x\in\Omega$，$(\mathrm{d}^kf)(x)\in L(\underbrace{\mathbf{R}^m,\cdots,\mathbf{R}^m}_{k\text{ 个因子}};\mathbf{R}^n)$.

3. 在 $\mathbf{R}^m$ 与 $\mathbf{R}^n$ 中均给出坐标系 $x$ 与 $y$ 以后，$f\in C^k$ 之充分必要条件是 $f$ 的各个分量均为古典意义下的具有直到 $k$ 阶在内的偏导数的函数.

4. 若 $f\in C^k$，则 $\mathrm{d}^kf(x)$ 对任意 $x\in\Omega$ 均为对称的多重线性映射. 用坐标来表示就是 $f$ 之各个分量的直到 $k$ 阶在内的偏导数之值，与施行构成这一偏导数之各个一阶偏导数之次序无关.

## §3 泰勒公式、莫尔斯引理、插值公式

**1. 泰勒公式** 在前面两节中我们已经指出，微分学的基本思想就是对一个映射 $f(x)$ 把

$(\Delta_h f)(x)=f(x+h)-f(x)$的线性部分分解出来，即给出以下的分解

$$f(x+h)-f(x)=A(x)h+o(h), \tag{1}$$

这里 $A$ 是一个线性映射．因为对不同点 $x$ 这个映射并不相同，所以它是一个函数(其值是线性映射或称线性算子，所以它是一个算子值函数)．如果这样的分解是可能的，就说 $f$ 在 $x$ 点可微，而 $A(x)$称为其微分，它也时常记为$(T_x f)(x)$或$(\mathrm{d}f)(x)$．在古典的微积分教本中则说 $Ah$ 才是微分，而且记 $h=\mathrm{d}x$，这样就有了古典的微分定义

$$\mathrm{d}f=f'(x)\mathrm{d}x.$$

用(1)式来作为微分的定义对变量 1 维，$f$ 也是 1 维——即我们常说的一元函数——是适用的：上式就是用于这个情况．因为这时 $A$ 是 $\mathbf{R}^1$ 到 $\mathbf{R}^1$ 的映射，因此是 1 维矩阵即一个数．我们就称为 $f$ 在 $x$ 之导数，记作 $f'(x)$．同时(1)式也适用于 $x$ 是 $m$ 维向量，即 $x=(x_1,\cdots,x_m)\in\Omega\subset\mathbf{R}^m$，而 $f$ 为 $n$ 维向量的情况．这时 $f:\Omega\to\mathbf{R}^n$ 在一定坐标系下应表为

$$y_i=f_i(x_1,\cdots,x_m), \qquad i=1,2,\cdots,n. \tag{2}$$

这时 $A$ 是一个线性映射 $\mathbf{R}^m\to\mathbf{R}^n$，不过这个 $\mathbf{R}^m$ 并不是$\Omega$ 所在的 $\mathbf{R}^m$，而是其切空间．后一个 $\mathbf{R}^n$ 则是靶空间 $\mathbf{R}^n$ 的切空间．如用上述坐标 $x,y$ 表示，则 $A$ 是下述矩阵

$$A=\frac{\partial(f_1,\cdots,f_n)}{\partial(x_1,\cdots,x_m)}. \tag{3}$$

(3)称为映射(2)的雅可比矩阵，这个映射现在称为 $f$ 的切映射．总之，我们在上一节的处理是把古典的微积分中一元函数和多元函数的微分统一在一起了．这样做还有一个目的，即将来这一个理论还可以适用于更广泛的领域中．例如 $\mathbf{R}^m$ 与 $\mathbf{R}^n$ 将被代以弯曲的空间，或代以无限维空间．

现在这样看待(1)式：由它立刻可见，$f(x)$在 $x$ 点是连续的．当 $\|h\|\to 0$ 时，(1)式左方是无穷小量．所以(1)表示这个无穷小量可以分解成线性部分和高阶部分(对于 $h$ 而言)．我们也说过，$(\Delta_h f)(x)$的线性部分包含了有关 $f$ 的性态的信息．所以，用 $\mathrm{d}f$ 代替 $\Delta_h f$ 称为$f$ 的线性化．我们暂时把如何通过线性化来研究 $f(x)$放在§5 以后再讲．现在要问，可否把 $o(h)$继续分解为 $h$ 的二次式而与$h^2$ 同阶，再加上 $h$ 的三次式而与$h^3$ 同阶……的无穷小量？上节结束时讲到高阶微分时已经提到这一点了．现在我们要继续沿这条思路走下去．

在上节讲到 $k$ 阶微分时，我们假设$f\in C^k$．现在因为所需的微分阶数不定，所以我们假设$f\in C^\infty$．

先看 $m=n=1$ 的情况．这时，我们有熟知的拉格朗日公式

$$f(x+h)-f(x)=f'(x+\theta h)h, \qquad 0<\theta<1.$$

其成立的条件只是 $f$ 在开区间$(x,x+h)$中有导数 $f'(x)$存在．如果 $f'(x)$在 $x$ 点还是连续可微的，则在上式右方对 $f'(x+\theta h)$再用一次拉格朗日公式有

$$f'(x+\theta h)=f'(x)+f''(x)\theta h+o(h).$$

代入前式则有

$$f(x+h)-f(x)=f'(x)h+f''(x)\theta h^2+o(h^2).$$

我们前面说过，在 $m=n=1$ 时，$f'(x)=A$，所以上式与(1)是一致的，而有

$$Ah+o(h)=f'(x)h+f''(x)\theta h^2+o(h^2),$$

也就是

$$o(h)=f''(x)\theta h^2+o(h^2).$$

但是它并不是我们需要的进一步的分解，因为其右方的 $\theta$ 并不确定．我们只知道 $\theta$ 依赖于 $h$，但具体的依赖方式并不知道，所以 $\theta h^2$ 并不是 $h$ 的二次式．回忆一下我们在上一节一开始时对于微分 $\mathrm{d}f$ 作为差分的"主要部分"的说法所作的评论（可以把 $o(h)$ 的一部分也归入 $Ah$ 而不损害它作为"主要部分"，而当 $A=0$ 时，$Ah=0$ 怎样也不能算是 $\Delta_h f$ 的主要部分），就知道我们提出问题时为什么说要把 $o(h)$ 分解成一个 $h$ 的二次式而与 $h^2$ 同阶（其实 $h$ 的二次式不一定必与 $h^2$ 同阶，因为有系数全为 0 的特例与上述 $A=0$ 相应）．可见要解决 $o(h)$ 的进一步分解还要用其他方法．

仍是为简单起见，下面令 $n=1$，也就是在(2)中把 $f$ 的各个分量分开来讨论，但是却令 $m>1$．只有在这时我们才会看到，由于缺少了拉格朗日公式带来的麻烦．如果我们是想在 $x$ 点附近研究 $f(x)$，则要求 $\Omega$ 是 $x$ 的这样一个邻域：若一点 $\tilde{x}\in\Omega$，则连接 $x$ 与 $\tilde{x}$ 的线段也全在 $\Omega$ 内．这类区域称为对 $x$ 为星形的区域，于是若 $x$ 与 $x+h$ 都在 $\Omega$ 中，则连接这两点的线段即集合

$$\{x+th, 0\leqslant t\leqslant 1\} \tag{4}$$

也全在 $\Omega$ 中，$f(x)$ 在此集合上的限制

$$F(t)=f(x+th)=f(x_1+th_1,\cdots,x_m+th_m)$$

成为 $t$ 的函数，而且因为已设 $f(x)\in C^\infty$，所以 $F(t)\in C^\infty([0,1])$，尤其值得注意的是

$$\begin{cases} F'(t)=\sum_{i=1}^{m}\dfrac{\partial f(x+th)}{\partial x_i}h_i, \qquad F'(0)=f'(x)h=(\mathrm{d}f)(x)h, \\ \cdots\cdots\cdots\cdots \\ F^{(k)}(t)=\sum\dfrac{\partial^k f(x+th)}{\partial x_{i_1}\cdots\partial x_{i_k}}h_{i_1}\cdots h_{i_k}, \\ F^{(k)}(0)=\sum\dfrac{\partial^k f(x)}{\partial x_{i_1}\cdots\partial x_{i_k}}h_{i_1}\cdots h_{i_k}=((\mathrm{d}^k f)(x);h,\cdots,h). \end{cases} \tag{5}$$

最后一个式子很值得注意，其左方是 $F(t)$ 的古典的 $k$ 阶导数在 $t=0$ 时的值，而右方则是上节最后讲的 $k$ 阶微分所成的 $k$ 次型（不是 $k$ 阶多重线性形式）在 $(h_{i_1},\cdots,h_{i_k})$ 处之值，$((\mathrm{d}^k f)(x);h,\cdots,h)$ 是 $k$ 次型的标准记号．

对于这个 $F(t)$，可以应用微积分的基本定理而有

$$\begin{aligned} F(1)-F(0)=f(x+h)-f(x)&=\int_0^1 F'(t)\mathrm{d}t \\ &=\sum_{i=1}^{m}h_i\int_0^1\frac{\partial f(x+th)}{\partial x_i}\mathrm{d}t. \end{aligned} \tag{6}$$

于是，$(\Delta_h f)(x)$ 的分解就归结为(6)式右方这些积分（$h$ 是其中的参数）对 $h$ 的分解．这一分解可以用分部积分法来实现．于是我们对 $\int_0^1 F'(t)\mathrm{d}t$ 应用分部积分法，而有

$$\begin{aligned} \int_0^1 F'(t)\mathrm{d}t &= -(1-t)F'(t)\Big|_0^1+\int_0^1(1-t)F''(t)\mathrm{d}t \\ &=F'(0)+\int_0^1(1-t)F''(t)\mathrm{d}t \\ &=\sum_{i=1}^{m}\frac{\partial f(x)}{\partial x_i}h_i+\int_0^1(1-t)F''(t)\mathrm{d}t \end{aligned}$$

$$=((\mathrm{d}f)(x),h)+\int_0^1(1-t)F''(t)\mathrm{d}t. \tag{7}$$

这里为什么要“积出”一个$(1-t)$而不用

$$\int_0^1 F'(t)\mathrm{d}t = tF'(t)\Big|_0^1-\int_0^1 tF''(t)\mathrm{d}t?$$

因为这样一来，右方第一项成为 $F'(1)=\sum\limits_{i=1}^m\frac{\partial f(x+h)}{\partial x_i}h_i=((\mathrm{d}f)(x+h),h)$这样我们不是将$f(x+h)$用$f(x),f'(x),\cdots$来展开了，而是将$f(x)$用$f(x+h),f'(x+h),\cdots$来展开.而且右方第二项再作下去就会正负相间.总之会得到许多麻烦.又,(7)式右方第一项即线性映射$(\mathrm{d}f)(x)$作用在向量$h$上之值.本来是记作$(\mathrm{d}f)(x)h$的，现在写作$((\mathrm{d}f)(x),h)$，是为了与下文统一.

(7)式右方第二项是

$$\int_0^1(1-t)F''(t)\mathrm{d}t=\sum_{i,j=1}^m\int_0^1\frac{\partial^2 f(x+th)}{\partial x_i\partial x_j}h_ih_j(1-t)\mathrm{d}t=\sum_{i,j=1}^m F_{ij}(x,h)h_ih_j,$$

这里

$$F_{ij}(x,h)=\int_0^1(1-t)\frac{\partial^2 f(x+th)}{\partial x_i\partial x_j}\mathrm{d}t$$

确实依赖于$h$，所以不能说(7)式右方第二项就是$h$的二次式.要想得到所需的分解，应对这一项再作分部积分.于是有

$$\begin{aligned}\int_0^1(1-t)F''(t)\mathrm{d}t &= -\frac{1}{2}(1-t)^2F''(t)\Big|_0^1+\frac{1}{2}\int_0^1(1-t)^2F^{(3)}(t)\mathrm{d}t\\ &=\frac{1}{2}F''(0)+\frac{1}{2}\int_0^1(1-t)^2F^{(3)}(t)\mathrm{d}t\\ &=\frac{1}{2}\sum_{i,j=1}^m\frac{\partial^2 f(x)}{\partial x_i\partial x_j}h_ih_j+\frac{1}{2}\int_0^1(1-t)^2F^{(3)}(t)\mathrm{d}t.\end{aligned}$$

现在可以和前面的$o(h)=f''(x)\theta h^2+o(h^2)$来比较了.如果把我们现在的推算用于它(即$m=n=1$的情况).将得到

$$o(h)=\frac{1}{2}f''(x)h^2+\frac{1}{2}\int_0^1(1-t)^2F^{(3)}(t)\mathrm{d}t.$$

注意到由(5)式，$F^{(3)}(t)=O(t)h^3$，代入上式立即有

$$f''(x)\theta h^2+o(h^2)=\frac{1}{2}f''(x)h^2+O(1)h^3.$$

因此当$h\to0$时，至少在$f''(x)\neq0$处有$\theta\to\frac{1}{2}$即是说，拉格朗日公式中的$\theta$确切依赖于$h$的状况虽然还不知道，至少可以得出$\theta=\frac{1}{2}+o(1)$.

这样的计算当然还可以继续作下去，即对积分项逐次应用分部积分法，有

$$\begin{aligned}\frac{1}{2}\int_0^1(1-t)^2F^{(3)}(t)\mathrm{d}t &= -\frac{1}{3!}(1-t)^3F^{(3)}(t)\Big|_0^1+\frac{1}{3!}\int_0^1(1-t)^3F^{(4)}(t)\mathrm{d}t\\ &=\cdots\end{aligned}$$

$$= \sum_{k=3}^{N} \frac{1}{k!}[\mathrm{d}^k f(x); h, \cdots, h] + \frac{1}{N!}\int_0^1 (1-t)^N F^{(N+1)}(t). \tag{8}$$

在$[(\mathrm{d}^k f)(x); h, \cdots, h]$中是把$(\mathrm{d}^k f)(x)$作为一个 $k$ 线性映射而作用到$k$ 个$h$ 上去.

我们再来分析一下这个证明.我们假设了 $\Omega$ 对$x$ 是星形域,这个假设的作用在于,也仅仅在于它保证了连接 $x$ 与$x+h$ 的线段全在$\Omega$ 内,而沿着这样一个线段可以作积分,拉格朗日公式那样的 $\theta$ 也就可以从$\int_0^1 (1-t)^k F^{(k+1)}(t)\mathrm{d}t$ 中对$F^{(k+1)}(t)$这个因子应用积分中值定理得出.但这只对 $n=1$ 时可用.(就是应用以下定理:若 $f(x)$,$g(x)$,均为连续函数,而且 $g(x) \geqslant 0$,则$\int_a^b f(x)g(x)\mathrm{d}x = f(c)\int_a^b g(x)\mathrm{d}x$,$a \leqslant c \leqslant b$.但当 $n>1$ 时,$F^{(k+1)}(t)$是一个 $n$ 维向量,而不是一个函数,所以这个定理不能用.)即令按这样讲,我们现在对 $f$ 的假设仍然多于拉格朗日公式$f(b)-f(a)=f'(c)(b-a)$.在那里只要求 $f(x)$在$[a,b]$上连续,$(a,b)$上有导数,至于 $f'(x)$是否连续,因此$\int_a^b f'(x)\mathrm{d}x$ 是否有意义尚且不知,更不说它是否等于 $f(b)-f(a)$,更不说能否分部积分了.初学微积分的读者常认为:既然微分与积分互为逆运算,则$\int_a^b f'(x)\mathrm{d}x = f(b)-f(a)$自然成立.这样是把问题看得太简单了.我们在第四章将用很大的力量还给不出一个完满的答案!暂时把它放在一边,我们现在加上星形域这个条件,就在一个十分重要的方面与拉格朗日公式所需的假设一致了.如果不问这公式中所需要的可微性条件,则这个公式所受到的真正的限制在于:它只能用于一个区间上,而即令是在 $\mathbf{R}^1$ 中,区间还只是最简单的开集.现在举一个"反例":令 $I=(0,1)\cup(2,3)$,$I$ 自然是一个开集.今在 $I=[0,1]\cup[2,3]$上定义函数 $f(x)$如下

$$f(x) = \begin{cases} 0, & x \in [0,1], \\ 1, & x \in [2,3], \end{cases}$$

则 $f(x)$该是一个规矩的函数了,可是令 $a=0$,$b=3$,立刻可见拉格朗日公式不成立,区间与一般开集之区别至少在于区间对其每个内点都是星形的!当然这个"反例"简单得令人好笑.可是到了高维空间,星形假设就不那么好笑了.我们前面证明"当$(\mathrm{d}f)(x)=0$ 时,$f(x) \equiv$常数时也遇到这个问题.星形的本质就是 $\Omega$ 可以"缩"为其中一点 $x$,而 $\Omega$ 在收缩、变形过程中没有越出 $\Omega$,$x$ 则可以完全不动.收缩、变形都是数学中极为重要的概念,我们在§2 中两次讲到连续拓展法就是与此密切相关的.我们证明§2 的定理 9 时,也就是利用这个方法,引入了一个连续变动的参数 $t$,才得到 $f(x+h)-f(x)-(\mathrm{d}f)(x)h$ 的精确估计.

现在把这个问题放在一边.我们把$(\Delta_h f)(x)$分解为 $h$ 的各次"幂",目的是为了在 $x$ 附近研究$f(x)$.即令 $\Omega$ 对于$x$ 不是星形区域,$x$ 既是$\Omega$ 之内点,必有某个以 $x$ 为心,以充分小的正数 $\rho$ 为半径的球 $B_\rho(x) \subset \Omega$.球对于球内的点是星形区域,所以只要以 $B_\rho(x)$代替 $\Omega$,总可以证明相应的结果在 $x$ 附近成立.因此,下面我们就不再提星形区域的问题,而所得结果也只说在 $x$ 附近成立.上面用了很大篇幅只是因为这是一个重要方法,而以后我们还会多次使用它.将来还会把它抽象化为某种类型数学模式.不过它已超出本书范围之外了.总结起来,我们有

**定理 1** 若 $f: \Omega \to \mathbf{R}^n$ 是$C^\infty$ 映射,这里 $\Omega \subset \mathbf{R}^m$ 是一开集,$x \in \Omega$,则对充分小的 $h$ 以及任意正整数$N$ 必有

$$f(x+h)=f(x)+\sum_{k=1}^{N}\frac{1}{k!}[\mathrm{d}^k f(x);\underbrace{h,\cdots h}_{k\text{个}}]+R_N(x,h),\tag{9}$$

这里

$$R_N(x,h)=\frac{1}{N!}\int_0^1(1-t)^N F^{(N+1)}(t)\mathrm{d}t,\tag{10}$$

$F(t)=f(x+th)$. (9)称为 $f(x)$在 $x$ 点的$N$ 阶泰勒公式,$R_N(x,h)$称为其余项,而且

$$\|R_N(x,h)\|=O(1)\|h\|^{N+1}.\tag{11}$$

证　除(11)以外都已证明了. (11)也是很容易的. 因为 $F^{(N+1)}(t)=\sum\limits_{|\alpha|=N+1}(\partial_x^\alpha f)(x+th)h^\alpha$. 尽管$\partial_x^\alpha f(x+th)$中还含有 $h$,它不是$h$ 的 $N+1$ 次多项式,但是它在 $x$ 的某个闭邻域中(注意$\|h\|$充分小)为 $C^\infty$ 函数,从而有界,所以$\|F^{(N+1)}(t)\|\leqslant M\|h\|^{N+1}$. 代入(10)式即得证明. 这里的记号$\partial_x^\alpha$ 下面再解释.

**注 1**　泰勒公式有多种形式的余项,(10)称为积分余项,重要的是估计式(11).

**注 2**　我们现在换一个记号,在(9)式中把 $x$ 换成 0 而把 $h$ 换成$x$,将有

$$f(x)=f(0)+\sum_{k=1}^{N}\frac{1}{k!}[(\mathrm{d}^k f)(0);x,\cdots,x]+R_N(0,x),\tag{12}$$

这里$[(\mathrm{d}^k f)(0);x,\cdots,x]$既然是 $k$ 线性映射作用在$k$ 个向量$(x,\cdots,x)$上所得之值,自然是 $x$ 之各个分量$(x_1,\cdots,x_m)$的 $k$ 次齐次多项式. 在数学中有一套记号特别适用于表示 $m$ 个变元(例如这里的$(x_1,\cdots x_m)$)的运算. 首先我们引入重指标 $\alpha=(\alpha_1,\cdots,\alpha_m)$,这里每个 $\alpha_i$ 都是非负整数,$\alpha_i\in\mathbf{N}$,$\mathbf{N}$ 是自然数的集合(我们都习惯于称 1,2,…为自然数,而 0 不算自然数,但在比较新的数学文献中常把 0 也算作自然数的,于是 $\mathbf{N}=\{0,1,2,\cdots\}$,我们现采用这种记号),所以也记作 $\alpha\in\mathbf{N}^m$. 我们记$|\alpha|=\alpha_1+\cdots+\alpha_m\geqslant 0$,称为 $\alpha$ 之长度. $\alpha>0$ 即指每一个 $\alpha_i>0$,若有两个重指标 $\alpha=(\alpha_1,\cdots,\alpha_m)$与 $\beta=(\beta_1,\cdots,\beta_m)$,$\alpha\pm\beta=(\alpha_1\pm\beta_1,\cdots,\alpha_m\pm\beta_m)$. 当然这里应要求 $\alpha_i\pm\beta_i\geqslant 0$,否则重指标$(\alpha_1\pm\beta_1,\cdots,\alpha_m\pm\beta_m)$就没有意义了. $\alpha\geqslant\beta$ 就表示 $\alpha_i\geqslant\beta_i$, $i=1,2,\cdots,m$. $\alpha!=\alpha_1!\cdots\alpha_m!$. 值得注意的是二项系数的记号也推广到了重指标上:$\binom{\alpha}{\beta}=\frac{\alpha!}{\beta!\ (\alpha-\beta)!}=\binom{\alpha_1}{\beta_1}\cdots\binom{\alpha_m}{\beta_m}$. 这当然只在 $\alpha\geqslant\beta$ 时有意义,若不然必有例如 $\alpha_1<\beta_1$,这时按惯例应定义$\binom{\alpha_1}{\beta_1}=0$. 根据这个惯例,我们也应该规定,若 $\alpha\geqslant\beta$ 不成立则$\binom{\alpha}{\beta}=0$. 此外对于 $x$ 我们定义 $x^\alpha=x_1^{\alpha_1}\cdots x_m^{\alpha_m}$. 我们又定义$\partial_x^\alpha=\partial_{x_1}^{\alpha_1}\cdots\partial_{x_m}^{\alpha_m}$. 这样的定义仍能保持例如 $x^\alpha\cdot x^\beta=x^{\alpha+\beta}$,$\partial_x^\alpha\cdot\partial_x^\beta=\partial_x^{\alpha+\beta}$ 等关系式成立.

还是回到(12)式,现在它成为

$$f(x)=f(0)+\sum_{1\leqslant|\alpha|\leqslant N}A_\alpha x^\alpha+R_N(0,x).$$

而且 $R_N(0,x)=O(1)\|x\|^{N+1}$. 对上式双方求导数,立即有

$$\partial_x^\alpha f(0)=\alpha!\ A_\alpha.$$

因此 $A_\alpha=\frac{1}{\alpha!}\partial_x^\alpha f(0)$,如果再注意到 $0!=1$,$\partial_x^0 f=f$,则上式还可以写成

$$f(x)=\sum_{|\alpha|\leqslant N}\frac{1}{\alpha!}\alpha_x^\partial f(0)x^\alpha+R_N(0,x).\tag{13}$$

这样得到的泰勒公式形式与一元函数的泰勒公式完全相同不过我们要注意，$\sum\limits_{|\alpha|\leqslant N}$ 表示对一切长度不大于 $N$ 的重指标求和.

**注 3**　由推导出泰勒公式的推理过程，可以得到一个十分有用的结果如下：

**引理 2**（**阿达玛**（J. Hadamard）**引理**）　设 $f(x)$ 在 $x=0$ 的某个邻域 $\Omega$ 中为 $C^k$ 映射，$\Omega\subset\mathbf{R}^n$，且 $f(0)=0$，这时必存在 $C^{k-1}$ 映射 $g_1,\cdots,g_m:\Omega\to\mathbf{R}^n$，使得

$$f(x)=\sum_{i=1}^m x_i g_i(x).\tag{14}$$

**证**　利用前面关于星形区域的说明，我们不妨设 $\Omega$ 关于 0 是星形区域. 于是对 $x\in\Omega$，有含于 $\Omega$ 内的线段 $I:\{tx:0\leqslant t\leqslant 1\}$ 连接 0 与 $x$ 两点. 考虑到 $f$ 在此直线段 $I$ 上的限制为 $t$ 的函数：

$$F(t)=f\Big|_I=f(tx).$$

显然 $F(t)$ 对 $t$ 是 $k$ 阶连续可微的，而有

$$F(1)-F(0)=f(x)-f(0)=\int_0^1 F'(t)\mathrm{d}t=\sum_{i=1}^m x_i\int_0^1\frac{\partial f}{\partial x_i}(tx)\mathrm{d}t.$$

因为 $f(0)=F(0)=0$. 又有

$$f(x)=\sum_{i=1}^m x_i\int_0^1\frac{\partial f}{\partial x_i}(tx)\mathrm{d}t.$$

令 $g_i(x)=\int_0^1\frac{\partial f}{\partial x_i}(tx)\mathrm{d}t$，显然 $g_i\in C^{k-1}(\Omega):\Omega\to\mathbf{R}^n$ 即合于所求.

**2. 极值问题**　微分学最重要的应用之一是求极值. 不仅是在 **17** 世纪费马的时代如此，而且从那时以来几个世纪中极大极小问题都是数学和数学物理中人们注意力的焦点之一. 它所采用的方法，所产生的新概念和新思想早已超越了微分学的领域. 我们这里仅只能就古典的情况，就多元函数的局部极值看一下微分学的应用.

在古典的极值问题中我们只考虑函数在某区域的内点取得的极值，而现代的许多极值问题，例如出现在规划问题中的极值，时常是在区域边界上取得的. 这一些与我们下面所讨论的很不一样. 我们在下面又不打算深究函数光滑性问题，所以我们假设 $f:\Omega\to\mathbf{R}$ 是一个 $C^2$ 函数，$\Omega$ 是一个开区域，还要注意 $f$ 的靶空间 $\mathbf{R}^n$ 就是 $\mathbf{R}$，即 $n=1$，因此 $f$ 就是一个函数，而不是一个高维向量：一个高维向量是谈不上什么极值问题的.

所谓极值是一个局部概念，准确一些说，我们有

**定义 1**　设 $f(x):\Omega\to\mathbf{R}$，$x_0\in\Omega$ 是 $f(x)$ 的一个局部极小（或者说 $f(x)$ 在 $x_0$ 达到局部极小），如果存在 $x_0$ 的一个邻域 $U$，使对一切 $x\in U$，都有

$$f(x)\geqslant f(x_0).$$

类似地可以定义局部极大. 以下简称为极小和极大.

我们处理这种局部极小（或极大）的方法，仍与前面证明泰勒公式的方法一样，不失一般性，取 $\Omega$ 为对 $x_0$ 星形的区域（例如 $B_\rho(x_0)$），则对充分小的 $h$，线段

$$I=\{x_0+th;0\leqslant t\leqslant 1\}$$

全在 $\Omega$ 内，并连接 $x_0$ 与 $x=x_0+h$ 两点，令 $F(t)=f(x_0+th)$，则 $F(t)$ 是 $t\in I$ 的函数．而且若设 $f\in C^k$，则 $F(t)\in C^k(I)$，且 $F(1)=f(x_0+h)=f(x)$，$F(0)=f(x_0)$．因此

$$\begin{aligned}F(1)-F(0)&=\int_0^1 F'(t)\mathrm{d}t=-(1-t)F'(t)\Big|_0^1+\int_0^1(1-t)F''(t)\mathrm{d}t\\&=F'(0)+\int_0^1(1-t)F''(t)\mathrm{d}t\\&=\sum_{i=1}^m\frac{\partial f(x_0)}{\partial x_i}h_i+\sum_{i,j=1}^m\int_0^1(1-t)\frac{\partial^2 f(x_0+th)}{\partial x_i\partial x_j}h_ih_j\mathrm{d}t.\end{aligned}$$

如果必要，可以把 $\Omega$ 向 $x_0$ 缩小一点（例如把 $B_\rho(x_0)$ 的半径缩小一点）而设 $\dfrac{\partial^2 f(x_0+th)}{\partial x_i\partial x_j}$ 在 $\overline{\Omega}$ 上连续．这样做的目的在于保证 $\dfrac{\partial^2 f(x_0+th)}{\partial x_i\partial x_j}$ 对于 $t$ 和 $h$ 为一致连续，因此

$$\frac{\partial^2 f(x_0+th)}{\partial x_i\partial x_j}=\frac{\partial^2 f(x_0)}{\partial x_i\partial x_j}+r(t,h).$$

这里 $r(t,h)$ 当 $\|h\|\to 0$ 时，对 $(t,h)$ 一致地趋于 0，因此，只要 $\|h\|<\delta(\varepsilon)$，必有 $|r(t,h)|<\varepsilon$．将上式代入前面的积分，就会得到我们所需的基本的估计式．

$$f(x)=f(x_0)+\sum_{i=1}^m\frac{\partial f(x_0)}{\partial x_i}h_i+\frac{1}{2!}\sum_{i,j=1}^m\frac{\partial^2 f(x_0)}{\partial x_i\partial x_j}h_ih_j+o(1)\|h\|^2. \tag{15}$$

现在设 $x_0$ 是局部极小，并一项一项地估计上式，先看 $h$ 的一次项，我们把过 $x_0+h$，$x_0$ 以及过 $x_0-h$，$x_0$ 的两个线段合并成一个．后一段不妨写成 $[x_0+(-t)h, x_0]$，因此合并以后得到一个较长的线段 $[x_0-h, x_0+h]$ 或 $\alpha(t)=\{x_0+th, -1\leqslant t\leqslant 1\}$．这样做的目的是使 $x_0$ 相应于 $[-1,1]$ 的内点 $t=0$．于是记 $F(t)=(f\circ\alpha)(t)$，$h$ 的一次项 $\sum\limits_{i=1}^m\dfrac{\partial f(x_0)}{\partial x_i}h_i=\mathrm{d}F(0)=\mathrm{d}f(x_0)\cdot\dot{\alpha}(0)$．由假设 $f$ 以 $x_0$ 为极小．所以 $F(t)$ 作为 $t$ 的函数，以 $[-1,1]$ 的内点 $t=0$ 为极小．因此，由费马定理 $\mathrm{d}F(0)=0$，即是说

$$\mathrm{d}f(x_0)\cdot\dot{\alpha}(0)=0,\qquad \forall\,\alpha(t).$$

但是前面已说过，每一条过 $x_0$ 的按接触意义的等价类就是一个向量．我们用直线 $\alpha(t)=\{x_0+th, -1\leqslant t\leqslant 1\}$ 作为它的代表元，就可以得到这个向量的坐标表示为 $h=(h_1,\cdots,h_m)$．$\mathrm{d}f(x_0)$ 作为由 $x_0$ 的切空间 $\mathbf{R}^m=\{h\}$ 到靶空间 $\mathbf{R}^1$ 的切空间（还是 $\mathbf{R}^1$）的线性映射，既然把所有的向量 $\dot{\alpha}(0)$ 都映为 0，自然是一个零映射：

$$\mathrm{d}f(x_0)=0.$$

如果读者感到这样的推理太抽象，不妨按通常的微积分教本中的讲法再讲一次：$f(x)$ 既以 $x_0$ 为极小，记 $x$ 之分量为 $(x_1,\cdots,x_m)$，$x_0$ 之分量为 $(x_0^{(1)},\cdots,x_0^{(m)})$，则在 $f(x)=f(x_1,\cdots,x_m)$ 中，我们可以依次地只让一个分量变化．于是可以考查 $f(x_1,x_0^{(2)},\cdots,x_0^{(m)})$，作一个一元函数，应在 $x_1=x_0^{(1)}$ 时达到极小．因此，由费马定理 $\dfrac{\partial f(x_1,x_0^{(2)},\cdots,x_0^{(m)})}{\partial x_1}=0$ 于 $x_1=x_0^{(1)}$ 时，亦即 $\dfrac{\partial f(x_0)}{\partial x_1}=0$．同理 $\dfrac{\partial f(x_0)}{\partial x_i}=0$，$i=2,\cdots,m$．这样，我们就会得到

$$\mathbf{grad} f(x_0) = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_m}\right)(x_0) = 0. \tag{16}$$

所谓只让 $x_1$ 变，就是在(15)中令 $h = (h_1, 0, \cdots, 0)$，所以$\dfrac{\partial f(x_0)}{\partial x_1} = 0$ 也就是

$$\frac{\partial f(x_0)}{\partial x_1}h_1 + \frac{\partial f(x_0)}{\partial x_2}0 + \cdots + \frac{\partial f(x_0)}{\partial x_m}0 = \sum_{i=1}^{m} \frac{\partial f(x_0)}{\partial x_i}h_i = 0,$$

这正是 $df(x_0) \cdot \dot{\alpha}(0) = 0$.

既然问题如此简单，又何必用与坐标无关的讲法让缺少经验的读者摸不着头脑呢？问题在于，我们会遇到需要在许多坐标系之间作变换的情况.因此，与坐标无关的处理是最方便的.我们正是想多通过一些例子帮助读者逐步熟悉这种处理问题的方法.

现在来看关于 $h$ 的二次项$\dfrac{1}{2}\sum\limits_{i,j=1}^{m}\dfrac{\partial^2 f(x_0)}{\partial x_i \partial x_j}h_i h_j$.它是一个二次型，其系数矩阵$\left(除了因子\dfrac{1}{2}\right)$即 $f$ 之黑塞矩阵，亦即 $(d^2 f)(x_0)$.记此矩阵为 $A$，则上节定理 8 已表明 $A$ 是一个对称矩阵，而这一项就是$\dfrac{1}{2}\langle Ah, h\rangle$.这里我们引用了线性代数中 $\mathbf{R}^m$ 中的内积记号 $\langle \cdot, \cdot \rangle$.线性代数告诉我们，任一个实对称矩阵都有实特征根 $\lambda_1, \cdots, \lambda_m$.不妨设 $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_m$(现在的函数 $f$ 自然是实值函数，否则谈不上极值)且必可用正交变换化为对角形，即

$$T^{-1}AT = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix}. \tag{17}$$

$\lambda_i$ 是 $A$ 的特征根.特征根也称为固有值.在 $\langle Ah, h\rangle$ 中令 $h = Tk$，$k$ 仍为 $\mathbf{R}^n$ 中的向量，则

$$\begin{aligned}\langle Ah, h\rangle &= \langle ATk, Tk\rangle = \langle {}^t TATk, k\rangle \\ &= \langle T^{-1}ATk, k\rangle.\end{aligned}$$

这些特征根有以下几种情况：

1. 所有 $\lambda_i > 0$.这时 $A$ 称为正定的(positive definite).而且

$$\langle Ah, h\rangle = \langle T^{-1}ATk, k\rangle \geqslant \lambda_1 \|k\|^2 = \lambda_1 \|h\|^2.$$

这里我们利用了正交变换不改变向量之长度：$\|h\| = \|Tk\| = \|k\|$.

2. 所有 $\lambda_i \geqslant 0$，这时 $A$ 称正半定(positive semi-definite)或非负定(non-negative definite)，而且对一切 $h \neq 0$ 有

$$\langle Ah, h\rangle \geqslant 0.$$

3. 所有 $\lambda_i \leqslant 0$.这时 $A$ 称为负半定(negative semi-definite)或非正定(non-positive definite)而且对一切 $h \neq 0$ 有

$$\langle Ah, h\rangle \leqslant 0.$$

4. 所有 $\lambda_i < 0$，这时 $A$ 称为负定(negative definite)，且

$$\langle Ah, h\rangle \leqslant -|\lambda_m| \|h\|^2 (= \lambda_m \|h\|^2).$$

注意，因为 $\lambda_m < 0$，所以 $\lambda_m = -|\lambda_m|$.

5. 有的 $\lambda_i$ 为正，有的为负，这时 $A$ 称为不定的(indefinlte).$\langle Ah, h\rangle$ 对某些 $h \neq 0$ 取正值(不

仅是非负值)，对某些 $h\neq 0$ 取负值(不仅是非正值).

关于 $A$ 的特征根或固有值的符号，只有这几个情况. 有了它们，$\frac{1}{2}[\mathrm{d}^2 f(x_0);h,h]=\frac{1}{2}\langle(\mathrm{d}^2 f)(x_0)h,h\rangle$ 的符号就完全确定了.

(15)的最后一项是余项. 对任意 $\varepsilon>0$ 必可找到一个 $\delta(\varepsilon)>0$，使当 $\|h\|<\delta$ 时，$|o(1)\|h\|^2|<\varepsilon\|h\|^2$. 但是就是这一项给我们带来麻烦.

所谓极值问题，其实就是当 $\|h\|$ 相当小时，希望 $f(x_0+h)-f(x_0)$ 的符号能保持不变. 首先我们看到 $(\mathrm{d}f)(x_0)$ 必须为 0. 因为(14)第二项是 $(\mathrm{d}f)(x_0)\cdot h$ 它其实是向量 $(\mathrm{d}f)(x_0)$ 与向量 $h$ 的"内积"($(\mathrm{d}f)(x_0)$ 与 $h$ 这两个向量有本质上的不同. 为什么其本质不同反而能定义其"内积"，将在第七章中讨论). 如果有一个 $h$ 使 $(\mathrm{d}f)(x_0)h>0$，则采用 $-h$ 必有 $(\mathrm{d}f)(x_0)(-h)<0$，因此 $f(x_0+h)-f(x_0)$ 与 $f(x_0-h)-f(x_0)$ 必反号(后面 $\|h\|$ 高次项不会影响于此. 读者仿照下文自会证出). 因此 $x_0$ 不可能是极值. 其实上面我们已经用费马定理证明了这一点. 问题是我们在讨论极值问题时，是在讨论一种特别简单的映射：$n=1$，因而 $f$ 就是一个普通的函数，$(\mathrm{d}f)(x_0)=0$ 就是 $\frac{\partial f(x_0)}{\partial x_i}=0, i=1,2,\cdots,m$，或写作 $\mathbf{grad}f(x_0)=0$. 在这种点上会出现极值等非常有趣的现象. 当 $n>1$ 时，是否要 $f$ 之一切分量之一阶偏导数均在 $x_0$ 处为 0，在 $x_0$ 处才出现"奇迹"呢？(我们前面说过，"奇点"就是出现"奇迹"的点.)用不着. 注意现在 $(\mathrm{d}f)(x_0)$ 是一个 $1\times m$ 矩阵，其秩最多是 1. $(\mathrm{d}f)(x_0)=0$ 就是 $(\mathrm{d}f)(x_0)$ 之秩不是最大的 1，而是更小的 0！推广这一点到一个一般的映射 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$. 我们说，凡 $(\mathrm{d}f)(x)$ 不能取最大秩的点 $x_0$ 就是奇点. 在现在的情况称为临界点(critical point). 非临界点称为正规点(regular point). 在§5中会看到，正规点都是规规矩矩的点. 若 $y_0$ 是函数在其临界点 $x_0$ 的值：$y_0=f(x_0)$，$y_0$ 就称为临界值(critical value). 不过临界值也可能是 $f$ 在某正规点 $x_1$ 之值：$y_0=f(x_1)$. 这是因为临界值的原像可能有许多个，例如这里的 $x_0$ 和 $x_1$，而不一定它们都是临界点.

有了这些根本概念，我们就可以得到下面的

**定理 3** 若 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}$ 在 $\Omega$ 中为 $C^2$ 映射，$x_0$ 是 $\Omega$ 的内点，则 $f$ 以 $x_0$ 为极小(极大)的必要条件是

1. $(\mathrm{d}f)(x_0)=0$，

2. $(\mathrm{d}^2 f)(x_0)$ 为正半定(负半定).

**证** 1. 不需再证.

2. $(\mathrm{d}^2 f)(x_0)$ 不是正半定(正定也包括在其内)就是有一个负特征根，设为 $\lambda_1$. 又设相应的特征向量，又称固有向量(eigenvector)是 $h^{(1)}$. $\|h^{(1)}\|\neq 0$，但其值不定，因为特征向量再乘上一个常数因子仍是特征向量. 于是我们考虑 $f(x_0+h^{(1)})-f(x_0)$. 因为 $(\mathrm{d}f)(x_0)=0$，所以(15)之第二项为 0. 至于第三项，由特征根与特征向量之定义.

$$(\mathrm{d}^2 f)(x_0)h^{(1)}=\lambda_1 h^{(1)},$$

所以，记 $h^{(1)}=(h_1^{(1)},\cdots,h_m^{(1)})$，有

$$\frac{1}{2}\sum_{i,j=1}^{m}\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}h_j^{(1)}h_j^{(1)}=\frac{1}{2}[(\mathrm{d}^2 f)(x_0);h^{(1)},h^{(1)}]$$
$$=\frac{1}{2}\langle(\mathrm{d}^2 f)(x_0)h^{(1)},h^{(1)}\rangle=\frac{\lambda_1}{2}\|h_1\|^2.$$

最后余项中的 $o(1)$ 当 $\|h\|$ 充分小时可以任意小，因此取 $\|h^{(1)}\|$ 充分小，使 $|o(1)|<\frac{|\lambda_1|}{4}$，或者说 $o(1)<-\frac{\lambda_1}{4}$（注意 $\lambda_1<0$）. 把这些结果代回(14)即知，当 $\|h^{(1)}\|$ 充分小时

$$f(x_0+h^{(1)})-f(x_0)\leqslant\frac{\lambda_1}{4}\|h^{(1)}\|^2<0.$$

这与 $x_0$ 是 $f(x)$ 之极小显然是矛盾的.

关于 $x_0$ 是极大值的情况证明类似，定理证毕.

**定理 4** 若 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}$ 在 $\Omega$ 中为 $C^2$ 映射，$x_0$ 是 $\Omega$ 之内点，则 $f$ 以 $x_0$ 为极小（极大）的充分条件是

1. $(\mathrm{d}f)(x_0)=0$，
2. $(\mathrm{d}^2 f)(x_0)$ 为正定（负定）.

**证** 当 1 成立时，我们有，对任意 $h\in\mathbf{R}^m$

$$f(x_0+h)-f(x_0)=\frac{1}{2}\sum_{i,j=1}^{m}\frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}h_i h_j+o(1)\|h\|^2$$
$$=\frac{1}{2}\langle(\mathrm{d}^2 f)(x_0)h,h\rangle+o(1)\|h\|^2.$$

如果 $(\mathrm{d}^2 f)(x_0)$ 为正定，则其最小特征根 $\lambda_1>0$，如前面所述，当 $\|h\|$ 充分小时，余项为 $|o(1)|\|h\|^2<\frac{\lambda_1}{4}\|h\|^2$，故 $o(1)\|h\|^2>-\frac{\lambda_1}{4}\|h\|^2$.

$$f(x_0+h)-f(x_0)\geqslant\frac{\lambda_1}{2}\|h\|^2-\frac{\lambda_1}{4}\|h\|^2=\frac{\lambda_1}{4}\|h\|^2\geqslant 0.$$

所以 $x_0$ 是极小.

关于极大的情况证明类似.

大家可以看到，必要条件与充分条件之间仍有一些间隙. $(\mathrm{d}^2 f)(x_0)$ 为正半定表明作为矩阵，它可能有零特征根. 一个矩阵有零特征根的充分必要条件是其行列式为 0，在这个情况下这个矩阵或这个映射或这个临界点称为蜕化的. 在我们的情况下，如果用坐标系来表示，$(\mathrm{d}^2 f)(x_0)$ 就是黑塞矩阵 $\mathrm{Hess}_x f(x_0)=\left(\frac{\partial^2 f(x_0)}{\partial x_i \partial x_i}\right)$. 当它蜕化时，我们就说映射 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}$ 的临界点 $x_0$ 是蜕化的. 因此，$x_0$ 为极小的必要条件与充分条件的间隙就在于：当 $x_0$ 为极小时，$(\mathrm{d}^2 f)(x_0)$ 可能是正定的，也可能是蜕化的，而得不出 $(\mathrm{d}^2 f)(x_0)$ 为正定. 反之，若设 $(\mathrm{d}^2 f)(x_0)$ 为正半定，则它可能是正定的（如果我们把正定也算作正半定的话），这时 $x_0$ 确定是极小；但它也可能是蜕化的，这时我们对 $f$ 在 $x_0$ 附近的性态就说不出所以然来了.

于是，余下的就只有 $(\mathrm{d}^2 f)(x_0)$ 为不定的. 这时，它既可能是蜕化的，又可能是非蜕化的. 但

不论如何，其特征根有正有负，不妨设 $\lambda_1>0,\lambda_2<0$，而相应的特征向量为 $h^{(1)},h^{(2)}$. 于是仿照前面的作法很容易看到，当 $\|h\|$ 充分小时

$$f(x_0+h^{(1)})-f(x_0)\geqslant\frac{\lambda_1}{4}\|h^{(1)}\|^2>0,\quad 若\ h\neq0;$$

$$f(x_0+h^{(2)})-f(x_0)\leqslant\frac{\lambda_2}{4}\|h^{(2)}\|^2<0,\quad 若\ h\neq0;$$

就是说，在某些方向上，$f(x_0)$很像极小，而在另一些方向上，$f(x_0)$又像极大. 这样的点称为鞍点(saddle point). 图 3-3-1 上画的是

$$z=f(x,y)=x^2-y^2,$$

这时

$$\mathrm{d}^2f(x,0)=2\begin{pmatrix}1&0\\0&-1\end{pmatrix}.$$

它有两个特征根 $\lambda_1,\lambda_2=1,-1$. 相应的特征向量是 $h^{(1)}=(1,0)$，$h^{(2)}=(0,1)$. $h^{(1)}$ 是 $x$ 轴方向的向量. 沿此方向 $f(x)$有极小. $h^{(2)}$ 是 $y$ 轴方向的向量，沿此方向 $f(x)$有极大. 但整个看来，$(0,0)$是一个鞍点. 但是$(0,0)$是映射(17)的一个非蜕化的临界点.

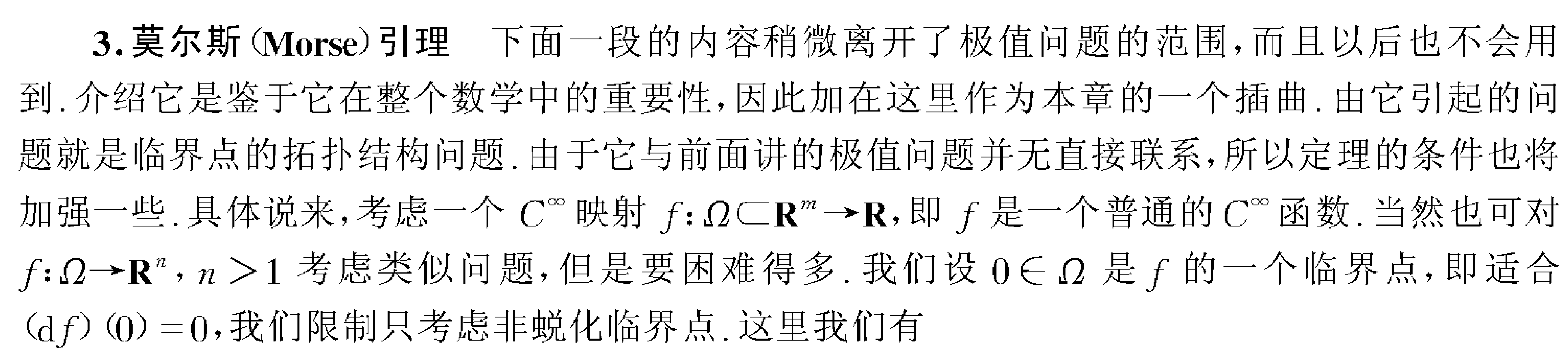


图 3-3-1

临界点的研究是现代数学中一个很活跃的分支——奇点理论的根本问题之一. 在这里划分临界点为蜕化与非蜕化是极为重要的. 蜕化临界点情况太复杂，而非蜕化临界点则情况完全清楚了，因为我们有重要的莫尔斯引理(Morse lemma).

**3. 莫尔斯(Morse)引理** 下面一段的内容稍微离开了极值问题的范围，而且以后也不会用到. 介绍它是鉴于它在整个数学中的重要性，因此加在这里作为本章的一个插曲. 由它引起的问题就是临界点的拓扑结构问题. 由于它与前面讲的极值问题并无直接联系，所以定理的条件也将加强一些. 具体说来，考虑一个 $C^\infty$ 映射 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}$，即 $f$ 是一个普通的 $C^\infty$ 函数. 当然也可对 $f:\Omega\to\mathbf{R}^n$，$n>1$ 考虑类似问题，但是要困难得多. 我们设 $0\in\Omega$ 是 $f$ 的一个临界点，即适合 $(\mathrm{d}f)(0)=0$，我们限制只考虑非蜕化临界点. 这里我们有

**定义 2** 若对上述 $C^\infty$ 映射 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}$，$0\in\Omega$ 是 $f$ 之临界点，即 $(\mathrm{d}f)(0)=0$，而且 $(\mathrm{d}^2f)(x)$ 是非奇异的，如用坐标表示，即设 $f$ 在 $x=0$ 处的黑塞矩阵 $\mathrm{Hess}_xf(0)$ 适合 $\det(\mathrm{Hess}_xf(0))\neq0$，则 0 称为 $f$ 之非蜕化临界点.

这里我们的基本结果是

**定理 5(莫尔斯引理)** 若 0 是上述 $f$ 的非蜕化临界点，则有 0 的一个充分小的邻域以及在此邻域中的局部坐标 $y$，$x=x(y)$，使 $0=x(0)$，而且 $f$ 可以化为标准形式(或称法式：normal form)：

$$f(x)=f[x(y)]=\Phi(y)=f(0)+\sum_{i=1}^{\lambda}y_i^2-\sum_{j=1}^{m-\lambda}y_{\lambda+j}^2. \tag{18}$$

**证** 不失一般性可以设 $f(0)=0$. 所谓局部坐标就是一个微分同胚 $x=x(y)$，亦即对应 $y\to x$ 为一一映射，且 $x(y)$与其反函数 $y=y(x)$在 0 点附近均属于 $C^\infty$.

这个定理与线性代数中二次型必可用非奇异的线性变换化为标准形式的定理是非常相近的. 它有多种证法. 下面我们介绍米尔诺(J. Milnor)的一个证明，此证明也十分相近于线性代数中关于二次型的定理的证明. 证明引自米尔诺著《莫尔斯理论》一书 6～8 页(科学出版社，1988).

由阿达玛引理(即引理 2),一定有

$$f(x)=f(0)+\sum_{j=1}^{m}x_jg_j(x)=\sum_{j=1}^{m}x_jg_j(x),$$

而且有

$$g_i(0)=\frac{\partial f(0)}{\partial x_i}=0.$$

于是再用一次阿达玛引理:

$$g_i(x)=\sum_{k=1}^{m}x_kh_{ki}(x),$$

且 $h_{ki}(0)=\dfrac{\partial g_i(0)}{\partial x_k}$. 以它代入 $f(x)$ 的表达式即得

$$f(x)=\sum_{k,j=1}^{m}x_kx_jh_{kj}(x). \tag{19}$$

我们可以假设

$$h_{kj}(x)=h_{jk}(x).$$

因为取上式中的两项,并将其和改写为

$$\begin{aligned}x_kx_jh_{kj}(x)+x_jx_kh_{jk}(x)&=x_kx_j\frac{1}{2}[h_{kj}(x)+h_{jk}(x)]+x_jx_k\frac{1}{2}[h_{kj}(x)+h_{jk}(x)]\\&=x_kx_j\bar{h}_{kj}(x)+x_jx_k\bar{h}_{jk}(x),\end{aligned}$$

显然 $\bar{h}_{jk}=\bar{h}_{kj}$. 把这两项代回 $f(x)$ 之表达式. 即知 $h_{kj}(x)=h_{jk}(x)$. 而且易见 $h_{kj}(0)=\dfrac{\partial^2 f(0)}{\partial x_j\partial x_k}$. 所以 $\det(h_{kj}(0))=\det(\text{Hess}_x f(0))\neq 0$.

$\left(\dfrac{\partial^2 f(0)}{\partial x_j\partial x_k}\right)$ 的主对角线上必有非 0 元. 因为,如果 $h_{11}=0$,则(19)中至少有一个 $h_{1i}\neq 0$. 否则 $f(x)$ 的表达式中将不出现因子 $x_1x_i\,(\forall\, i)$. 所以 $\dfrac{\partial^2 f(0)}{\partial x_1\partial x_i}=0$ 而 $\text{Hess}_x f(0)$ 是蜕化的. 这样我们就可以设 $h_{11}=0$,但 $h_{12}\neq 0$. 这时,再令 $x_1=y_1+y_2$, $x_2=y_1-y_2$,则 $f(x)$ 中将出现

$$h_{12}x_1x_2+h_{21}x_2x_1=2h_{12}(y_1^2-y_2^2).$$

此外再也不会有含 $y_1^2$ 和 $y_2^2$ 的因子了. 所以在作了这样的变量变换以后,$f(x)$ 的表达式中 $y_1^2$ 的系数为 $2h_{12}\neq 0$. 但这就是在新变量下的 $h_{11}$,所以我们说可设(19)中 $h_{11}\neq 0$.

将(19)重写为

$$\begin{aligned}f(x)&=h_{11}x_1^2+\left(\sum_{i=2}^{m}h_{1i}x_ix_1+\sum_{i=2}^{m}h_{i1}x_1x_i\right)+\cdots\\&=h_{11}x_1^2+2\left(\sum_{i=2}^{m}h_{1i}x_i\right)x_1+\cdots\\&=\pm(\sqrt{|h_{11}|}\,x_1)^2+2\,\frac{\sum\limits_{i=2}^{m}h_{1i}x_i}{\sqrt{|h_{11}|}}\,(\sqrt{|h_{11}|})\,x_1+\cdots.\end{aligned}$$

这里 $\pm$ 表示 $h_{11}$ 之符号,即 $h_{11}=\pm|h_{11}|$,“$\cdots$”中不含因子 $x_1$,经过配方,有

$$f(x)=\pm\left(\sqrt{|h_{11}|}x_1\pm\sum_{i=2}^{m}h_{1i}x_i\Big/\sqrt{|h_{11}|}\right)^2+\sum_{i,j>1}\bar{R}_{ij}(x)x_ix_j.$$

如果作变量变换

$$y_1=\sqrt{|h_{11}|}x_1\pm\sum_{i=2}^{m}h_{1i}x_i\Big/\sqrt{|h_{11}|},$$
$$y_i=x_i,i>1, \tag{20}$$

则

$$f(x)=\pm y_1^2+\sum_{i,j>1}\bar{h}_{ij}(x)y_iy_j. \tag{21}$$

问题在于(20)是否微分同胚.将在§5中证明的反函数定理表明,只需(20)的雅可比矩阵在 $x=0$ 处可逆即可.但是这个雅可比矩阵是

$$\begin{pmatrix}\sqrt{|h_{11}|}+O(x) & * & \cdots & * \\ 0 & 1 & & 0 \\ & & \ddots & \\ 0 & 0 & & 1\end{pmatrix},$$

其中 $*$ 表示某个在 $x=0$ 处光滑的函数,因此在 $x=0$ 处这个矩阵自然是可逆的.所以在 $x=0$ 的一个充分小邻域中,(20)是一个微分同胚,从而(21)是 $y$ 在 $y=0$ 附近的 $C^\infty$ 函数,而(21)中的 $\bar{h}_{ij}(x)$ 其实也是 $y$ 的 $C^\infty$ 函数.若记(21)为

$$f(x)=\pm y_1^2+\bar{f}(y)=\pm y_1^2+\sum_{i,j=2}^{m}\bar{h}_{ij}(x)y_iy_j,$$

在 $\bar{f}$ 中视 $y_1$ 为参数,很容易证明 $(\mathrm{d}\bar{f})(0)=0$(这里 $(\mathrm{d}\bar{f})(0)$ 表示对 $y_2,\cdots,y_{m-1},y_m$ 的微分,而 $y_1$ 作为参数也为0).我们也容易看到

$$\mathrm{Hess}_x f(0)=\begin{pmatrix}\pm 2 & 0 & \cdots & 0\\ 0 & & & \\ & & \mathrm{Hess}_y\bar{f}(0) & \\ 0 & & & \end{pmatrix}.$$

$\mathrm{Hess}_y\bar{f}(0)$ 表示只对 $y_2,\cdots,y_m$ 求导,但在参数 $y_1=0$ 以及 $y_2=\cdots=y_m=0$ 时求值.因为 $\mathrm{Hess}_x f(0)$ 是非奇异的,所以 $\mathrm{Hess}_y\bar{f}(0)$ 也是非奇异的.这样,可以对 $\bar{f}(y)$ 重复上面的作法(但以 $y_1$ 为参数),经有限次以后,即是(18)式.定理证毕.

有了莫尔斯引理后我们换一个角度来看一个映射 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$ 在 $0\in\Omega$ 处的泰勒公式

$$f(x)=f(0)+\sum_{1\leqslant|\alpha|=k}^{N}A_\alpha x^\alpha+R_N(0,x). \tag{22}$$

上一节我们讲过两条曲线 $x=f_1(t)$ 和 $x=f_2(t)$ 在 $t=0$ 处 $k$ 阶接触的概念,即

$$f_1(t)-f_2(t)=o(1)|t|^k,\text{但右方不能改为 } o(1)|t|^{k+1}.$$

曲线是 $I\subset\mathbf{R}\to\mathbf{R}^n$ 的映射,即(22)中 $m=1$ 的情况,这个概念当然可以推广到两个映射 $f_1,f_2$: $\Omega\subset\mathbf{R}^m\to\mathbf{R}^n,0\in\Omega$.我们说这两个映射在 $x=0$ 处 $N$ 阶接触,即指.

$$f_1(x)-f_2(x)=o(1)\|x\|^N,\text{但右方不能改为 } o(1)\|x\|^{N+1}.$$

所以(22)就表明 $f(x)$ 与 $f(0)+\sum_{1\leqslant|\alpha|=k}^{N}A_\alpha x^\alpha\ (A_\alpha=\frac{1}{\alpha!}\partial_x^\alpha\dot{f}(0))$ 在 $x=0$ 处 $N$ 阶接触.泰勒公式的这一部分,即这一“段”,称为其 $N$ 阶节(jet).这是一个 $N$ 次多项式,但是要注意适合 $|\alpha|=k$ 的 $A_\alpha x^\alpha$ 有很多项,这是由于 $\alpha$ 是一个重指标的原故.所谓 $N$ 阶节是指要把所有适合 $|\alpha|=N$ 的 $A_\alpha x^\alpha$ 都算进去,而不能有遗漏.只有这样才能保证接触的阶数是 $N$.如果我们把所有适合 $|\alpha|\leqslant N-1$ 的项算进去了,而适合 $|\alpha|=N$ 的项漏了一项,则所得并不是 $N$ 阶节,而实质上与 $N-1$ 阶节一样.这就是说 $N$ 阶节的定义还应该表述得更清晰一些,这一点我们就不讲了.所以泰勒公式也可以这样来陈述:即 $f(x)$ 之 $N$ 阶节与 $f(x)$ 有 $N$ 阶接触.当然我们可以证明下面的命题:若有一个 $N$ 次多项式 $P_N(x)$ 与 $f(x)$ 在 $x=0$ 处有 $N$ 阶接触,则 $P_N(x)$ 必是 $f(x)$ 的 $N$ 阶节.即由 $f(x)-P_N(x)=o(1)\|x\|^N$,必可证明 $P_N(x)$ 是 $f(x)$ 的 $N$ 阶节.其实证明很容易,令 $J_N(x)=f(0)+\sum_{1\leqslant|\alpha|=k}^{N}\frac{1}{\alpha!}\partial^\alpha f(0)x^\alpha$ 为 $f(x)$ 之 $N$ 阶节,则由 $f(x)-J_N(x)=o(1)\|x\|^N$ 知

$$P_N(x)-J_N(x)=[P_N(x)-f(x)]+[f(x)-J_N(x)]=o(1)\|x\|^N.$$

但是两个 $N$ 次多项式若不恒等,其差最多是 $O(1)\|x\|^N$,而不是 $o(1)\|x\|^N$.因此 $P_N(x)=J_N(x)$.

以上说的是 $N$ 阶节与 $f(x)$ 之误差比之 $\|x\|^N$ 为更高阶的无穷小.但是我们还要进一步问,$f(x)$ 与其 $N$ 阶节是否可能在 $x=0$ 附近从拓扑学角度看来是等价的?所谓“从拓扑学角度来看是等价的”这句话并不明确.现在我们可以这样理解它:如果在 $x=0$ 附近有一个微分同胚 $x=x(y)$ 使 $0=x(0)$,则 $f(x)=f[x(y)]=F(y)$.这时我们就说 $f$ 和 $F$ 从拓扑学角度看来是等价的.

对一般的映射 $f:\Omega\subset\mathbf{R}^m\rightarrow\mathbf{R}^n$,讨论上述意义下的等价性是很困难的.但是莫尔斯引理告诉我们,若 $n=1$,而 $0\in\Omega$ 是 $f$ 的非蜕化临界点,则在 $x=0$ 附近,$f(x)$ 与一个二次型在拓扑学上是等价的.如果我们换用莫尔斯引理的一个更细微的证明,还可以进一步证明 $f(x)$ 与它的 2 阶节 $f(0)+\frac{1}{2}\sum_{i,j=1}^{m}\frac{\partial^2 f(0)}{\partial x_i\partial x_j}x_ix_j$ 是等价的.这些都是很深刻的结果.

**4. 插值公式** 迄今为止,我们讨论微分学问题时都是遵循下面的路径:先考虑离散的量,例如一段有限长的时间 $\Delta t$,一段有限长的距离 $\Delta x$,然后再考虑它的极限状况:$\Delta t\rightarrow 0,\Delta x\rightarrow 0$ 又如何如何.可以说,我们走的都是由离散到连续这样一条路.而且实践告诉我们,这是研究种种物理现象有力的工具.牛顿创立的微积分之所以如此重要,这是根源所在.然而可以进一步提出一个更深入的问题,即表现一个自然现象的“曲线”等等,其本身果然是连续的、光滑的……吗?例如,为了表示一段时间内温度的变化,我们可以用记录仪器自动地画出一条曲线来,而且我们感到有十足的把握说:温度作为时间的函数就是这条曲线.但是,许多记录仪器给我们的曲线,其实只是一些分布很密的点:粗看起来确实是像一条连续的曲线,细看起来更像一条虚线.而且不论是粗看还是细看,都还只是自然现象(通过观测和记录的仪器)作用于我们的感官的结果,而不是自然现象的本身.要想掌握自然现象的本身,除了需要越来越精密的仪器以外,还需要深刻的理论思考.但是我们的感觉,不论用多么好的仪器来加强,用多么深刻的理论来深化,也只可能是对自然现象的近似的描述.所以,上面说的由离散到连续之路,也只是我们的思维所走过的、越来越深刻地通向大自然的本质的道路.但是我们既然接受了从连续的角度来描述自然现象,当然也应该接

受从离散的角度来描述自然现象.

回顾一下我们已讲过的内容,就可以看到离散的东西与连续的东西是彼此对立而又互相补充的.微分学的最重要的定理之一的拉格朗日公式,告诉我们,不但$\frac{\Delta f}{\Delta x}$之极限是导数$f'(x)$,而且其本身也就是导数$f'(\xi)$,只不过$\xi$这个点的位置不能确定而已.其实泰勒公式也是一样,只不过把函数改变量$\Delta f(x)$(它是离散的),改用泰勒公式的节(少一项$f(0)$)来近似,其余项也有不确定处.所以泰勒公式也可认为是拉格朗日公式的推广.再从泰勒公式的节的几何意义来看更可以看到这一点.从$m=1,n=1$,即$f$表示一条曲线来看.它的1阶节$f(0)+f'(0)x$表示切线,但我们可以通过曲线上两点$P$、$Q$:$(0,y_0)$与$(h,y_1)$作割线

$$y-f(0)=\frac{y_1-y_0}{h}(x-0),$$

其极限(当$Q\to P$时)就是切线.所以1阶节是某一"离散量"的极限.若通过曲线上三点$P$、$Q$、$R$作一条二次曲线$y=Ax^2+Bx+C$,则当$Q$、$R\to P$时其极限位置也恰好是二阶节(若过这三点作一个圆,则其极限位置仍是一个圆,称为$P$点处的密切圆(osculating circle)或其曲率圆,其半径为曲率半径,恰好表征曲线在$P$点处的弯曲程度).仿此,若在此曲线上取$N+1$个点$(x_i, f(x_i))$,$i=1,2,\cdots,N+1$,并通过它们作一条$N$次曲线$y=A_0x^N+\cdots+A_N$,则当这$N+1$个点都趋向一点时,这个$N$次曲线似乎应该趋向于此曲线在该点的$N$阶节,从而保证达到最佳的接触阶数.而这个$N$次多项式应该能很好地描述$f(x)$在该点附近的性态.但是为此首先应该解决如下的问题,即插值问题:设给定$N+1$个离散的点:$(x_i,y_i)$,$i=1,\cdots,N+1$,怎样在一个指定的函数类中找到一个函数$y=f(x)$使得$y_i=f(x_i)$.或者用物理的语言来陈述它,设在某个实验或观测中发现,对应于某一物理量$x_1,\cdots,x_{N+1}$,另一物理量的相应值是$y_1,\cdots,y_{N+1}$,如何在一个指定的函数类中找到一个函数,使其图像通过这$N+1$个点$(x_i,y_i)$.有了这样一个函数自然十分有利于研究其中的物理规律性.上面我们都是讲的在一个指定的函数类中去找这个$f(x)$.函数类的限制是十分重要的.这种函数类应该是更容易处理的函数,因为"真正的"(应该说是"更准确的")函数关系是很难找到或者很难处理的.最简单的情况下可以取多项式,或三角多项式等等.因此最简单的插值问题如下:设在$x$轴上有$N+1$个点,依次排列为$x_1<x_2<\cdots<x_{N+1}$,找一个$N$次多项式$P_N(x)$,使$P_N(x_i)$等于$N+1$个指定的值$P_N(x_i)=y_i$.

这个问题很容易回答.因为拉格朗日插值公式给出的

$$\begin{aligned}P_N(x)&=y_1\frac{(x-x_2)\cdots(x-x_{N+1})}{(x_1-x_2)\cdots(x_1-x_{N+1})}+y_2\frac{(x-x_1)(x-x_3)\cdots(x-x_{N+1})}{(x_2-x_1)(x_2-x_3)\cdots(x_2-x_{N+1})}\\&\quad+\cdots+y_{N+1}\frac{(x-x_1)\cdots(x-x_N)}{(x_{N+1}-x_1)\cdots(x_{N+1}-x_N)}\\&=\sum_{j=1}^{N+1}y_j\prod_{i\neq j}\frac{(x-x_i)}{(x_j-x_i)}\end{aligned}\tag{23}$$

就是其解.这里每一项的分子都是一个$N$次多项式,而由$N$个形如$x-x_i$因子的组成:第一项缺少$x-x_1$第二项缺少$x-x_2$,$\cdots$,第$N+1$项缺少$x-x_{N+1}$.每一项的分母恰好是把分子中的$x$换成了缺失了的$x_j$.最后再乘上"系数"$y_j$.条件中$x_1,x_2,\cdots,x_{N+1}$的次序并无关系,但是不允许有$x_i=x_j$,否则上式会有分母为零的项.这种情况下,拉格朗日公式问题的提法需做相应的修

正.

如果 $y_i$ 是某一函数 $f(x)$ 在 $x_i$ 处之值：$y_i = f(x_i)$，则上式给出用一个多项式去逼近 $f(x)$ 的公式.

如果这些给定的 $x_i$（称为结点）是等距的，即令 $x_1 = a$，$x_2 = x_1 + \Delta x = a + \Delta x$，$x_3 = x_2 + \Delta x = a + 2\Delta x$，…，$x_{N+1} = x_N + \Delta x = \cdots = a + N\Delta x$，将得到拉格朗日插值公式的一个特例：牛顿插值公式

$$Q_N(x) = f(a) + \frac{x-a}{1!}\frac{\Delta f(a)}{\Delta x} + \frac{(x-a)(x-a-\Delta x)}{2!}\frac{\Delta^2 f(a)}{\Delta x^2} + \cdots + \frac{(x-a)\cdots(x-a-N\Delta x)}{N!}\frac{\Delta^N f(a)}{\Delta x^N}. \tag{24}$$

这里我们应用了差分记号：$\Delta f(a)$ 表示 $f(x)$ 在 $x = a$ 处对 $x$ 的增量 $\Delta x$ 的相应增量

$$\Delta f(a) = f(a + \Delta x) - f(a).$$

$\Delta f(a)$ 的增量记为 $\Delta(\Delta f(a)) = \Delta^2 f(a)$，称为二阶差分，所以

$$\Delta^2 f(a) = \Delta f(a + \Delta x) - \Delta f(a).$$

但是

$$\begin{aligned}\Delta f(a+\Delta x) &= f[(a+\Delta x)+\Delta x] - f(a+\Delta x)\\ &= f(a+2\Delta x) - f(a+\Delta x),\end{aligned}$$

所以

$$\begin{aligned}\Delta^2 f(a) &= [f(a+2\Delta x) - f(a+\Delta x)] - [f(a+\Delta x) - f(a)]\\ &= f(a+2\Delta x) - 2f(a+\Delta x) + f(a).\end{aligned}$$

仿此有

$$\begin{aligned}\Delta^k f(a) &= f(a+k\Delta x) - \frac{k}{1!}f(a+(k-1)\Delta x) + \frac{k(k-1)}{2!}f(a+(k-2)\Delta x) + \cdots + (-1)^k f(a)\\ &= \sum_{l=0}^{k}(-1)^{k-l}C_k^l f(a+l\Delta x).\end{aligned} \tag{25}$$

由此也可以得出

$$\begin{aligned}f(a+\Delta x) &= f(a) + \Delta f(a),\\ f(a+2\Delta x) &= f(a+\Delta x) + \Delta f(a+\Delta x)\\ &= f(a) + 2\Delta f(a) + \Delta^2 f(a),\\ f(a+3\Delta x) &= f(a+2\Delta x) + \Delta f(a+2\Delta x)\\ &= f(a) + 3\Delta f(a) + 3\Delta^2 f(a) + \Delta^3 f(a),\\ &\cdots\cdots\end{aligned} \tag{26}$$

我们不讲如何由拉格朗日插值公式得出牛顿插值公式.但是很容易验证(24)的正确性.例如以 $x = a$ 代入(24)式后，(24)式就只余下第一项，故 $Q_N(a) = f(a)$.令 $x = a + \Delta x$ 就此余下前两项，从而 $Q_N(a+\Delta x) = f(a) + \dfrac{\Delta x}{1!}\dfrac{\Delta f(a)}{\Delta x} = f(a) + \Delta f(a) = f(a+\Delta x)$，余类推.

现在的问题是由(23)和(24)去逼近 $f(x)$ 的误差问题.这里首先要注意，既然在结点 $x_1$，$x_2$，…，$x_{N+1}$ 上 $P_N(x_k) = Q_N(x_k)$，则 $P_N(x) \equiv Q_N(x)$.这是因为 $P_N(x) - Q_N(x)$ 是一个 $N$ 次

多项式，而在 $N+1$ 个结点 $x_1, x_2, \cdots, x_{N+1}$ 上它又为 0. 除非它恒等于 0，是不会有 $N+1$ 个零点的. 因此下面我们就来估计 $f(x)-P_N(x)$，解决这个问题的基础是推广的罗尔定理：若 $f(x)$ 在 $[a,b]$ 上为 $N$ 阶可微，而且在 $N+1$ 个点 $a=x_1<x_2<\cdots<x_{N+1}=b$ 上为 0，则必可在 $(a,b)$ 中找到一点 $\xi$ 使 $F^{(N)}(\xi)=0$. 这个定理的证明很容易：只要对 $F(x_{i+1})=F(x_i)=0$ 应用罗尔定理，即知必有 $\xi_i\in(x_i, x_{i+1})$ 使 $F'(\xi_i)=0$. 再对 $F'(x)$ 在 $(\xi_i, \xi_{i+1})$ 上应用罗尔定理，仿此就可证明上述定理. 于是令

$$f(x)=P_N(x)+R_N(x),$$

立即有，在结点 $x_k=a+(k-1)\Delta x$ 上

$$R_N(x_k)=0, \qquad k=1,2,\cdots,N+1,$$

因此可以把 $R_N(x)$ 写成

$$R_N(x)=\frac{1}{(N+1)!}(x-x_1)\cdots(x-x_{N+1})\Psi(x).$$

现令

$$F(z)=f(z)-P_N(z)-\frac{1}{(N+1)!}(z-x_1)\cdots(z-x_{N+1})\Psi(x). \tag{27}$$

注意，最后一项中因子 $\Psi(x)$ 的自变量仍写为 $x$，其余的则改写成了 $z$，这样做的好处是，将(27)双方都对 $z$ 求 $N+1$ 阶导数时，有

$$F^{(N+1)}(z)=f^{(N+1)}(z)-\Psi(x). \tag{28}$$

这里我们利用了以下的事实：$P_N(z)$ 是 $z$ 的 $N$ 次多项式，故对 $z$ 求 $N+1$ 次导数后为 0. 现在对 $F(z)$ 作为 $z$ 的函数应用推广的罗尔定理. 由于 $F(x_i)=f(x_i)-P_N(x_i)=0, i=1,2,\cdots,N+1$. 同时在 $z=x$ 处它也为 0，所以在 $a$ 与 $x$ 之间一共有 $N+2$ 个零点，所以一定在 $a$ 与 $x$ 中存在一点 $\xi$，使 $F^{(N+1)}(\xi)=0$. 以此代入(28)即有

$$\Psi(x)=f^{(N+1)}(\xi).$$

代回(23)即得含余项的拉格朗日插值公式

$$f(x)=\sum_{j=1}^{N+1}f(x_j)\coprod_{i\neq j}\frac{(x-x_i)}{(x_j-x_i)}+\frac{f^{(N+1)}(\xi)}{(N+1)!}\coprod_{j=1}^{N+1}(x-x_j). \tag{29}$$

我们也可对 $Q_N(x)$ 应用上法. 因为上法中完全未涉及结点是否等距，于是又可得含余项的牛顿插值公式：

$$\begin{aligned}f(x)&=f(a)+\frac{x-a}{1!}\frac{\Delta f(a)}{\Delta x}\\&+\cdots+\frac{(x-a)(x-a-\Delta x)\cdots(x-a-(N-1)\Delta x)}{N!}\frac{\Delta^N f(a)}{\Delta x^N}\\&+\frac{1}{(N+1)!}(x-a)\cdots(x-a-N\Delta x)f^{(N+1)}(\xi).\end{aligned} \tag{30}$$

我们特别要注意(30). 令 $\Delta x\to 0$，则对固定的 $N$ 恒可设 $|N\Delta x|<1$，因此 $x_1=a, x_2=a+\Delta x_1, \cdots, x_{N+1}=a+N\Delta x$ 全在区间 $(a-1, a+1)$ 内，当 $\Delta x\to 0$ 时，式左与 $\Delta x$ 无关，极限自然是 $f(x)$，式右的前 $N+1$ 项可以逐项求极限，只余下最后一项的 $\xi$ 之位置是与 $N$ 和 $\Delta x$ 有关的. 它一定在 $a$ 与 $x$ 之间内，因此当 $\Delta x\to 0$ 时，也一定有一个收敛子序列，其极限仍记为 $\xi$，不过仍在

$a$ 与 $x$ 之间.这样我们又重新回到泰勒公式

$$f(x)=\sum_{k=0}^{N}\frac{f^{(k)}(a)}{k!}(x-a)^k+\frac{f^{(N+1)}(\xi)}{(N+1)!}(x-a)^{(N+1)}. \tag{31}$$

余项的形状与前面讲的积分形式不同,称为拉格朗日形式的余项.但当 $x-a\to0$ 时它仍是比 $(x-a)^N$ 更高阶的无穷小量.

值得注意的是,从历史上看泰勒(B. Taylor,1655—1731)在 1715 年正是用差分方法得出了牛顿插值公式(但没有余项).然后他大胆地令 $\Delta x=0$(应为 $\Delta x\to0$),令 $N=\infty$(应为 $N\to\infty$)而得出了泰勒级数.说他大胆是因为当时人们对无穷级数以及对极限过程都不清楚,也不太了解问题的严重性.但是泰勒就进入了我们将在下一节讨论的新领域.

# §4 解析函数与 $C^\infty$ 函数

**1. 幂级数** 泰勒大胆地从插值公式得出了以他命名的级数展开式——泰勒级数展开式:

$$f(x)=\sum_{k=0}^{\infty}\frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{1}$$

或者在多个变量即 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$ 而 $m,n\geqslant1$ 的情况下有

$$f(x)=\sum_{0\leqslant k=|\alpha|}^{\infty}\frac{f^{\alpha}(0)}{\alpha!}(x-a)^{\alpha}. \tag{2}$$

这里 $\alpha$ 是重指标.其实在那个时代什么是函数并没有划出清楚的界限.不但泰勒说不出他的结果是否适用于一切函数,甚至到了 18 世纪末,占主流的看法仍然认为所谓函数就是有限或无限的运算组合.而如(1)式这样的运算组合已经是够广泛的了.关于级数的收敛性,泰勒好像还没有这样一个概念.拉格朗日有一些初步的收敛性的概念,所以他是认为任意函数都可以写成(1)式.即令在今天,许多初学微积分的读者都会有一个错觉,以为只要 $f(x)$ 在 $a$ 附近为 $C^\infty$ 函数,则(1)式成立.这就大错特错了.一个古典的例子是考虑实变量 $x$ 在 $x=0$ 附近的函数

$$f(x)=\begin{cases}\mathrm{e}^{-\frac{1}{x}}, & x>0,\\ 0 & x\leqslant0.\end{cases}$$

用很简单的极限运算立即可知 $f(x)\in C^\infty$(包括 $x=0$),而且对一切非负整数 $k$,

$$f^{(k)}(0)=0.$$

因此,对它应用(1)式,并取 $a=0$,则(1)式左方不恒为 0 而右方恒为 0,可见(1)并非对一切 $C^\infty$ 函数都成立.因此,有必要将使得(1)式成立的函数专门划分为一类.这就是我们下面要讲的解析函数.

在给出解析函数定义以前,我们先要注意,(1)与(2)都是所谓幂级数——确切一些说是以 $x=a\in\mathbf{R}^m$ 为中心(center)的幂级数:

$$\sum_{n=0}^{\infty}a_n(x-a)^n \text{ 或 } \sum_{0\leqslant|\alpha|=k}^{\infty}a_\alpha(x-a)^\alpha.$$

后式中 $\alpha$ 是重指标.这时有一个重要的说明:

1. 正如在第二章中讲到指数函数与三角函数时指出的,当一个函数可以用幂级数表示时,

最好是立即进入复域.因此,在本节中讲到幂级数时,恒认为自变量 $x$,中心 $a$ 以及系数 $a_n$ 或 $a_\alpha$ 都是复数.

2. 在讨论自变量为复数的函数时,单复变量与多复变量的情况有重大的原则区别.因此,下面我们都只限于单自变量的函数 $f:\Omega\subset\mathbf{C}\to\mathbf{C}$ 的情况.$\mathbf{C}$ 指复平面.它具有复维数 1,但是也可看作实维数 2:$\mathbf{C}\simeq\mathbf{R}^2$.作为复自变量,我们总用 $z$ 表示,但 $z=x+\mathrm{i}y$,$x,y$ 表示 $z$ 的实部与虚部,都是实数.同样复的函数值总用 $w=u+\mathrm{i}v$ 表示.$u,v$ 是 $w$ 的实、虚部,都是实数.所以下面讨论的复函数(通常此词是指复自变量函数,另一个词是复值函数,则是指函数值 $w$ 为复值,而自变量则不一定)有两种表示法:

$$w=f(z)\qquad\text{或}\qquad u=u(x,y),v=v(x,y).$$

总之,本节中我们只讨论以下的幂级数.

$$w=\sum_{n=0}^{\infty}a_nz^n\quad\text{或}\quad w=\sum_{n=0}^{\infty}a_n(z-a)^n. \tag{3}$$

我们先从幂级数的一般性质的讨论开始.我们在第二章中已经说过,复变量函数的极限与连续性与复级数的收敛性、绝对收敛性及一致连续性等概念与基本的定理都与对应的实的情况是一样的.

幂级数的收敛性最基本的特点反映在下述阿贝尔定理中(我们只看(3)中 $a=0$ 的情况,一般情况相同).

**定理 1(阿贝尔(N. Abel)定理)** 对于级数(3),必可找到一个非负实数 $R$($R$ 允许取 $+\infty$ 值),使得

1) (3)在圆 $|z|<R$ 中绝对收敛,在 $|z|>R$ 处发散.

2) 若 $K\subset\{z;|z|<R\}$ 是任一紧集,则(3)在 $K$ 上一致收敛.

圆 $|z|<R$ 称为(8)之收敛圆,$R$ 称为收敛半径.

**注** 定理中指出 $R\geqslant0$,若 $R=0$,则 $|z|<R$ 没有意义,而说(3)之收敛半径为 0 就是指对一切 $|z|>0$,级数(3)均为发散.

**证** 1)的证明.设(3)在一点 $z_0\neq0$ 处收敛,令 $\rho=|z_0|>0$,今证(3)在 $|z|<\rho$ 处必绝对收敛.事实上,因 $\sum\limits_{n=0}^{\infty}a_nz_0^n$ 收敛,故其一般项必有界(实际上当 $n\to\infty$ 时趋于 0):$|a_nz_0|^n\leqslant M$.考虑级数 $\sum a_nz^n$ 其中 $z$ 适合不等式 $|z|<|z_0|=\rho$,我们有

$$|a_nz^n|=|a_nz_0^n|\left|\frac{z}{z_0}\right|^n\leqslant M\left|\frac{z}{z_0}\right|^n$$

但是 $\left|\dfrac{z}{z_0}\right|=\dfrac{|z|}{\rho}<1$,所以几何级数 $\sum\limits_{n=0}^{\infty}M\left|\dfrac{z}{z_0}\right|^n<+\infty$,由比较判别法即知 $\sum\limits_{n=0}^{\infty}|a_nz^n|<+\infty$,亦即(3)绝对收敛.

若(3)在一点 $z_0\neq0$ 处发散,则在 $|z|>|z_0|$ 处(3)亦必发散.事实上,若(3)在一点 $z_1$ 收敛,而 $|z_1|>|z_0|$,则把证明的前一部分用于 $\sum\limits_{n=0}^{\infty}a_nz_0^n$,即知它是绝对收敛的,而与(3)在 $z_0$ 处发散相矛盾.

现在看一个集合 $J=\{\rho>0$,(3)在 $|z|<\rho$ 处绝对收敛$\}$.有两个可能,一是 $J=\varnothing$,即是说

(3)当且仅当 $z=0$ 时收敛.这时我们取收敛半径 $R=0$.二是 $J$ 非空.先看 $J$ 非空的情况,令 $R=\sup J$,则 $R>0$.若 $z_0$ 点适合 $|z_0|<R$,现在要在 $J$ 中找一个数 $\rho$ 使 $|z_0|<\rho$($R$ 还未证明是 $J$ 中之数).记 $\varepsilon=R-|z_0|>0$,由于 $R$ 是 $J$ 之上确界,$J$ 中必有至少一个元 $\rho>R-\frac{\varepsilon}{2}$,而由定义知一定有:(3)在 $|z|<\rho$ 时绝对收敛.现在 $z_0$ 适合 $|z_0|=R-\varepsilon<R-\frac{\varepsilon}{2}=\rho$,所以(3)在 $z_0$ 处绝对收敛.即 $|z_0|<R$ 必导致(3)在 $z_0$ 处绝对收敛.所以实际上 $R$ 是 $J$ 中之数.又可以证明当 $|z_0|>R$ 时,(3)在 $z_0$ 处必发散.因为若(3)在 $z_0$ 处收敛,则令 $\varepsilon=|z_0|-R$,$\rho=|z_0|-\frac{\varepsilon}{2}=R+\frac{\varepsilon}{2}>R$,当 $|z|<\rho$ 时必有 $|z|<|z_0|-\frac{\varepsilon}{2}$,而(3)在 $z$ 处也绝对收敛.这样 $\rho\in J$ 而与 $R=\sup J$ 矛盾.总之,当 $J$ 非空时 $R$ 必是收敛半径而定理结论中的(1)成立.

若 $J$ 为空,则取 $R=0$,结论(1)平凡地成立.

2)的证明如下.若 $R=0$,则 $K=\varnothing$ 而结论平凡地成立.若 $R>0$,则 $\{z;|z|<R\}$ 是一开集,其紧子集 $K$ 必为有界的,而且 $K$ 必与 $\{z;|z|<R\}$ 之边界 $\{z;|z|=R\}$ 有正的距离,从而 $K$ 含于 $|z|\leqslant R-\frac{\delta}{2}$ 中,且 $\sum\limits_{n=0}^{\infty}\left|a_n\left(R-\frac{\delta}{2}\right)^n\right|$ 收敛.对于 $z\in K$,因为 $|a_nz^n|\leqslant\left|a_n\left(R-\frac{\delta}{2}\right)^n\right|$,由一致收敛性的 $M$ 判别法知道 $\sum\limits_{n=0}^{\infty}a_nz^n$ 在 $K$ 中一致收敛.

由阿贝尔定理得知幂级数一定有收敛半径 $R\geqslant 0$.下面给出 $R$ 一个具体表达式

**定理 2**(**阿达玛**) 幂级数(3)的收敛半径是

$$R=1\Big/\overline{\lim_{n\to\infty}}\sqrt[n]{|a_n|}.\tag{4}$$

**证** 证明分两部分,先证当 $|z|<R$ 时(3)必绝对收敛.首先,$\sqrt[n]{|a_n|}$ 的极限不一定存在,但上极限却是一定存在的,而且易见 $R\geqslant 0$.如果 $R=0$,则 $|z|<R$ 是一个空集而谈不上(3)是否在 $z$ 点收敛或绝对收敛.如果 $R>0$,则由上极限的定义,对于任意小的正数 $\varepsilon>0$,最多除有限多个 $n$,一切系数均适合

$$\sqrt[n]{|a_n|}\leqslant\frac{1}{R}+\varepsilon.$$

现在 $z$ 是固定的,故当 $|z|<R$ 时,一定可以找到一个实数 $\lambda:0<\lambda<1$,使 $|z|\leqslant\lambda R$.于是从某一个 $N_0$ 开始,当 $n>N_0$ 时

$$|a_nz^n|\leqslant \quad \frac{1}{\phantom{R}}$$

但是$\frac{2\lambda}{\lambda+1}<1$，所以由级数绝对收敛性的比较法则(它对于复的无穷级数仍成立，这一点读者可以自己验证)，即知(3)当$|z|<R$ 时绝对收敛.

证明的另一部分是求证当$|z|>R$ 时，级数(3)必发散.与上一部分不同，这一部分证明对 $R=0$ 也适用.由上极限的定义，对任一 $\varepsilon>0$，必有一个上升序列 $n_1<n_2<\cdots<n_k<\cdots\to+\infty$ 使$\sqrt[n_k]{|a_{n_k}|}\geqslant\frac{1}{R}-\varepsilon$，所以

$$|a_{n_k}z^{n_k}|\geqslant\left[\left(\frac{1}{R}-\varepsilon\right)|z|\right]^{n_k}.$$

但因 $z$ 是固定的且$|z|>R$，所以必可写出$|z|=R+\delta$ 而 $\delta$ 是一个固定的正数.适当取 $\varepsilon$ 使$\left(\frac{1}{R}-\varepsilon\right)(R+\delta)>1$.代入上式即有$|a_{n_k}z^{n_k}|\geqslant1$，即是说级数(3)的一般项不趋向 0 因而发散.注意若 $R=0$，上面用到$\frac{1}{R}$处都没有意义了.但 $R=0$ 即$\overline{\lim\limits_{n\to\infty}}\sqrt[n]{|a_n|}=+\infty$，因而对任意大的正数 $M$，必可找到子序列 $n_1<n_2<\cdots<n_k<\cdots$使$\sqrt[n_k]{|a_n|}\geqslant M-\varepsilon$.上面凡用到$\frac{1}{R}$处都换成$M$，即知论证仍然有效.故定理成立.

凡讲到幂级数(3)之收敛性时，总有两个特例.一类幂级数收敛半径 $R=0$，即(3)只在中心 $z=0$ 收敛.例如 $\sum\limits_{n=0}^{\infty}n!z^n$ 即是.但我们不必用阿达玛公式(4)来计算 $R$，因为那会涉及求$\overline{\lim\limits_{n\to\infty}}\sqrt[n]{|n!|}$.这个计算并不容易，而需要一个有名的，也非常有用的斯特林(Stirling)公式.第四章中我们还会遇到它.对于 $\sum\limits_{n=0}^{\infty}n!z^n$ 反而用比值$a_{n+1}/a_n$ 来求收敛半径要容易得多.另一类幂级数的收敛半径 $R=\infty$，即$\overline{\lim\limits_{n\to\infty}}\sqrt[n]{|a_n|}=0$(也就是$\lim\limits_{n\to\infty}\sqrt[n]{|a_n|}=0$，因为这时极限值一定存在).于是级数(3)对一切 $z$ 均收敛，这时(3)之和称为整函数.数学中许多重要的函数如 $e^z$，$\cos z$，$\sin z$……都是整函数.不论如何，幂级数在收敛圆内的性状已完全清楚.在收敛圆外也很简单:发散.但是在收敛圆周上情况却十分复杂.(3)可能收敛，可能绝对收敛，可能一致收敛一直到圆周……但也可能发散.总之，级数(3)在收敛圆周上的性状是一个专门的研究领域.

要注意，级数(3)在收敛圆内是绝对收敛的.无穷级数理论中有一个基本的事实:绝对收敛级数在代数运算性质上最接近于有限和:可以交换求和次序，适用分配律等等.因此，幂级数的运算性质最为简单.下面我们只举出几个最简单常用的性质而都不再证明.

首先是乘法.设有两个幂级数 $\sum\limits_{n=0}^{\infty}a_nz^n$ 与 $\sum\limits_{m=0}^{\infty}b_mz^m$，其收敛半径分别是 $R_1$ 与 $R_2$，则$\left(\sum\limits_{n=0}^{\infty}a_nz^n\right)\left(\sum\limits_{m=0}^{\infty}b_mz^m\right)$本来是一个二重级数 $\sum\limits_{m,n=0}^{\infty}a_mb_mz^{m+n}$.但是我们常用的是以下的写法，即按 $m+n=k$ 的大小来排列.即有

$$\left(\sum_{n=0}^{\infty}a_nz^n\right)\left(\sum_{m=0}^{\infty}b_mz^m\right)=\sum_{k=0}^{\infty}\left(\sum_{m+n=k}a_nb_m\right)z^k. \tag{5}$$

注意，内含的求和 $\sum\limits_{m+n=k}a_nb_m$ 只是有限的，而不存在收敛性问题，故(5)式不是一个二重级数.(5)

的收敛半径不小于 $\min(R_1,R_2)$.

其次是求函数的复合. 设有两个函数 $F(w)=\sum_{n=0}^{\infty}a_nw^n$，其收敛半径为 $R_1$，以及 $\psi(z)=\sum_{n=1}^{\infty}b_nz^n$，其收敛半径为 $R_2$，注意我们设 $b_0=0$，其理由从下面即可以看到. 现在我们要求 $(F\circ\psi)(z)$ 的幂级数展开式，一方面我们有

$$F[\psi(z)]=\sum_{n=0}^{\infty}a_n[\psi(z)]^n \tag{6}$$

但是，为使它收敛，就需要 $|\psi(z)|<R_1$. 因为我们已设 $b_0=0$，所以当 $z=0$ 时 $\psi(z)=0$，而当 $|z|$ 充分小时 $|\psi(z)|$ 也充分小. 我们假设 $b_0=0$ 的原因即在此. 这里我们用到了 $\sum_{n=1}^{\infty}b_nz^n$ 在它的收敛圆 $|z|<R_2$ 的任一紧子集上一致收敛，因而其和在 $|z|<R_2$ 内连续. 这里读者会问，为什么不说其和在 $|z|<R_2$ 之任一紧子集内连续？这是因为连续性是一点附近的性质. 任取一点 $z_0\in\{z;|z|<R_2\}$，则 $z_0$ 必可放在一个闭圆 $K$ 中，而 $K\subset\{z;|z|<R_2\}$. $K$ 是一个有界闭集即紧集，而 $\sum_{n=1}^{\infty}b_nz^n$ 在 $K$ 上一致收敛，从而其和 $\psi(z)$ 在 $K$ 上连续，当然也在 $z_0$ 连续. 因为 $z_0$ 是收敛圆内任一点，所以 $\psi(z)$ 在整个收敛圆内，而不只在其任一紧子集上连续. 但是一致连续性并不是对于某一点来说的，而是要指定一个区域——现在是收敛圆，而我们恰好不能证明(3)在收敛圆内一致收敛. 由 $\psi(z)$ 在 $|z|<R_2$ 内连续以及 $\psi(0)=0$ 即知一定有一个正数 $r>0$ 存在，使当 $|z|<r$ 时 $|\psi(z)|<R_1$ 从而级数(6)不但有有意义的各项，而且(6)收敛. 于是

$$F(\psi(z))=\sum_{n=0}^{\infty}a_n\left(\sum_{m=1}^{\infty}b_mz^m\right)^n. \tag{7}$$

但是我们可以利用上面讲的乘法来把 $\left(\sum_{m=1}^{\infty}b_mz^m\right)^n$ 写成幂级数

$$\left(\sum_{m=1}^{\infty}b_mz^m\right)^n=\sum_{k=n}^{\infty}B_{nk}z^k. \tag{8}$$

现在的问题是如何把它代入(7)式，为此要注意这个级数是由 $z^n$ 开始的，而且易见 $B_{nn}=b_1^n$. 因此当我们把(8)代入(7)，并且要求例如 $z^l$ 之系数时，则(7)中 $n>l$ 的各项中一定不会有 $z^l$ 出现. $z^l$ 只出现在相应于 $n=0,1,\cdots,l$ 各项中，而成为

$$\sum_{n=0}^{l}a_nB_{nl}=A_l，其中\ B_{00}=1.$$

我们只知道它有一个正的收敛半径，但是没有办法把它用 $R_1,R_2$ 显式地表示出来.

以上我们只提供了一个计算复合函数的幂级数的程序. 一般说来，它是相当繁冗的. 但是它是相当重要的. 许多很难的数学定理、数学公式都要靠这样的方法计算出来. 下面我们只举一个例子，即求函数的倒数.

设级数(3)有收敛半径 $R$，其和为 $\varphi(z)$. 我们设 $\varphi(0)=a_0\neq0$，现在来求 $1/\varphi(z)$. 为此我们把 $\varphi(z)$ 写成

$$\varphi(z)=a_0+a_1z+\cdots=a_0\left[1-\sum_{m=1}^{\infty}b_mz^m\right],$$

而 $b_m=-a_m/a_0$. 于是

$$\frac{1}{\varphi(z)}=\frac{1}{a_0}\frac{1}{1-\sum_{m=1}^{\infty}b_mz^m}.$$

由此可知，我们可以考虑

$$F(w)=\frac{1}{a_0}\frac{1}{1-w} \tag{9}$$

再把它与 $w=\sum_{m=1}^{\infty}b_mz^m$ 复合起来即可. 但是在整个"复分析"(即复变量函数的微积分学)中，"最重要"的幂级数展开式就是几何级数

$$\frac{1}{1-w}=1+w+w^2+\cdots=\sum_{n=0}^{\infty}w^n,$$

它的收敛半径是 1，这一点是所有读者在任何时刻都不可忘记的. 应用它和上面所说的，即有

$$\frac{1}{\varphi(z)}=\frac{1}{a_0}\sum_{n=0}^{\infty}\Big(\sum_{m=1}^{\infty}b_mz^m\Big)^n,$$

这里 $|z|$ 应充分小以保证 $\left|\sum_{m=1}^{\infty}b_mz^m\right|<1$. 但是我们得不到收敛半径的具体表达式.

更具体一点，我们来看 $\sec z=\frac{1}{\cos z}$ 的幂级数展开式. 现在因 $\cos z=1-\frac{z^2}{2!}+\frac{z^4}{4!}\cdots$，所以 $a_0=1$，而 $\sum_{m=1}^{\infty}b_mz^m=\sum_{m=1}^{\infty}\frac{(-1)^{m-1}}{2m!}z^{2m}$，所以

$$\begin{aligned}\sec z&=1+\Big(\sum_{m=1}^{\infty}\frac{(-1)^{m-1}}{2m!}z^{2m}\Big)+\Big(\sum_{m=1}^{\infty}\frac{(-1)^{m-1}}{2m!}z^{2m}\Big)^2+\cdots\\&=1+\frac{1}{2}z^2+\frac{5}{24}z^4+\cdots.\end{aligned}$$

它的系数有些什么规律性，几乎无法看出. 它的收敛半径是 $\pi/2$ 也是用其他方法得出的. 总之，这是许多困难而又深刻的数学定理的来源.

**2. 解析函数与泰勒级数** 级数的一致收敛性可用于解决级数所表示的函数的微分与积分问题. 所以首先要说明什么是复变量函数 $f(z)$ 在一点 $z_0$ 的导数. 这时的导数与微分等概念形式上均与实变量函数无异，但是实质上大不相同. 关于导数，我们仍定义

$$f'(z_0)=\lim_{h\to0}\frac{f(z_0+h)-f(z_0)}{h}=\lim_{z\to z_0}\frac{f(z)-f(z_0)}{z-z_0}. \tag{10}$$

这里 $h$ 是一个非 0 复数，$z=z_0+h$. 这样定义以后，实变量函数导数的基本性质，特别如莱布尼茨性质、复合函数导数的计算等都与实变量情况一样. $f$ 在 $z_0$ 点的微分仍定义为一个线性映射 $A:\mathbf{C}\to\mathbf{C}$ 使

$$f(z_0+h)-f(z_0)=Ah+o(h). \tag{11}$$

$A$ 就是"乘以 $f'(z_0)$". 在通常的教本上是把 $Ah$ 即 $\Delta f$ 之线性部分称为微分 $(\mathrm{d}f)(z_0)$ 的，而我们现在则用 $(\mathrm{d}f)(z_0)$ 表示一个线性映射，它作用在 $h$ 上的结果则是古典意义下的微分：

$$(\mathrm{d}f)(z_0)\cdot h=\text{古典意义下的微分}$$

$$= f'(z_0)\mathrm{d}z, \qquad \mathrm{d}z = h.$$

因为这一切与前面讲的没有区别，我们就不再重复. 问题是如何求 $f'(z_0)$. 由于至今我们还没有定义过多少具体的函数，如何计算导数就成了问题. 但有两个非常重要的导数完全可用初等方法算出. 一是对非负整数 $k$，$(z^k)' = kz^{k-1}$. 事实上，利用二项式定理，

$$\begin{aligned}(z^k)' &= \lim_{h\to 0}\frac{1}{h}[(z+h)^k - z^k] = \lim_{h\to 0}\left(kz^{k-1} + \frac{1}{2!}k(k-1)z^{k-2}h + \cdots\right)\\ &= kz^{k-1}.\end{aligned}$$

另一个是 $\left(\frac{1}{z}\right)' = -\frac{1}{z^2}$. 事实上

$$\begin{aligned}\left(\frac{1}{z}\right)' &= \lim_{h\to 0}\left(\frac{1}{z+h} - \frac{1}{z}\right)\Big/h = \lim_{h\to 0}\frac{-h}{z(z+h)}\cdot\frac{1}{h}\\ &= -\frac{1}{z^2}.\end{aligned}$$

此式自然只在 $z\neq 0$ 时成立. 类似他还有 $\left(\frac{1}{z^k}\right)' = -\frac{k}{z^{k+1}}$.

由此，如果允许对幂级数(3)逐项求导，则应有

$$f'(z) = \sum_{n=0}^{\infty} na_n z^{n-1}.$$

18 世纪中，拉格朗日就认为可以用这样的方法来定义任意函数的导数，而完全用不着无穷小等概念. 拉格朗日称此为纯粹"代数的方法". 这当然只是一个幻想，因为其中涉及无穷级数的逐项求导问题. 这是一个困难的问题，通常微积分教本中关于这个问题的结果也有点怪，例如说：设某区间 $I$ 上的级数 $\sum_{n=0}^{\infty} f_n(x)$ 在 $x_0 \in I$ 处收敛，且 $\sum_{n=0}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x}f_n(x)$ 在 $I$ 上一致收敛，则原级数必可逐项求导：$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=0}^{\infty} f_n(x) = \sum_{n=0}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x}f_n(x)$. 其所以如此是因为级数的逐项求导的问题被化成了 $\sum_{n=0}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x}f_n(x)$ 的逐项求积(实际上是求原函数). 但是对幂级数(3)，则定理 1 给了我们十分自然的结果：

**定理 3**　若幂级数(3)之收敛半径为 $R>0$，其和为 $f(z)$，则

$$f'(z) = \sum_{n=0}^{\infty} na_n z^{n-1}, \tag{12}$$

且收敛半径不变.

**证**　先证(12)之收敛半径仍为 $R$. 事实上

$$\begin{aligned}\overline{\lim_{n\to\infty}}\sqrt[n-1]{|na_n|} &= \lim_{n\to\infty}\sqrt[n-1]{n}\,\overline{\lim_{n\to\infty}}\sqrt[n-1]{|a_n|}\\ &= \lim_{n\to\infty}\sqrt[n-1]{n}\,\overline{\lim_{n\to\infty}}\left(\sqrt[n]{|a_n|}\right)^{\frac{n}{n-1}}.\end{aligned}$$

但是很容易看到 $\lim_{n\to\infty}\sqrt[n-1]{n} = 1$. 这是因为 $\sqrt[n-1]{n} > 1$ 故可写为

$$\sqrt[n-1]{n} = n^{\frac{1}{n-1}} = 1 + \alpha_n, \qquad \alpha_n > 0.$$

从而由二项式定理，

$$n=(1+\alpha_n)^{n-1}>1+\frac{1}{2}(n-1)(n-2)\alpha_n^2.$$

因此

$$0<\alpha_n^2<(n-1)\Big/\frac{1}{2}(n-1)(n-2)\to 0.$$

又利用求上极限问题可以化为求某个子序列的极限问题(我们不妨设$\{a_n\}$就是这个子序列)又知

$$\begin{aligned}\overline{\lim_{n\to\infty}}\left(\sqrt[n]{|a_n|}\right)^{\frac{n}{n-1}}&=\left(\lim_{n\to\infty}\sqrt[n]{|a_n|}\right)^{\lim\frac{n}{n-1}}\\&=\overline{\lim_{n\to\infty}}\sqrt[n]{|a_n|}=\frac{1}{R}.\end{aligned}$$

所以级数(12)之收敛半径不变.

再证(12)之和为$f'(z)$.设(12)之和为$f_1(z)$,它也定义在$|z|<R$上.又记(3)之部分和及余项分别为$S_N(z)$与$R_N(z)$:

$$S_N(z)=\sum_{n=0}^{N-1}a_nz^n.\ R_N(z)=\sum_{n=N}^{\infty}a_nz^n.$$

对收敛圆$|z|<R$内任一点$z_0$($|z_0|<R$是自然的,而且还可以找到正数$\rho$使$|z_0|<\rho<R$)考虑

$$\begin{aligned}&\frac{f(z)-f(z_0)}{z-z_0}-f_1(z_0)\\&=\left[\frac{S_N(z)-S_N(z_0)}{z-z_0}-S_N'(z_0)\right]+(S_N'(z_0)-f_1(z_0))+\frac{R_N(z)-R_N(z_0)}{z-z_0}.\end{aligned}$$

于是最后一项可以化为

$$\sum_{n=N}^{\infty}a_n(z^{n-1}+z^{n-2}z_0+\cdots+z_0^{n-1}).$$

而当$|z|<\rho$时有

$$\left|\frac{R_N(z)-R_N(z_0)}{z-z_0}\right|\leqslant\sum_{n=N}^{\infty}n|a_n|\rho^{n-1}.$$

但是从我们对收敛半径不变的证明中已可看到$\sum\limits_{n=0}^{\infty}n|a_n|\rho^{n-1}<+\infty$,因此,对任意$\varepsilon$,必可找到$N_0$,使当$n>N_0$时它的余项小于$\varepsilon/3$.即是说$n>N_0$时$\left|\dfrac{R_N(z)-R_N(z_0)}{z-z_0}\right|<\varepsilon/3$.这里$N_0$的选择与$\rho$有关,但与$z$无关,只要它适合$|z|<\rho$即可.再看第二项.因为(12)在$z=z_0$($z_0$在收敛圆中)处收敛,$f_1(z_0)$是(12)之和,$S_N'(z_0)$是它的部分和,所以如果有必要就把$N_0$放大一些,但仍然与$z$无关而只与$\rho$有关,使第二项绝对值也小于$\varepsilon/3$.现在我们把$N_0$固定,并开始考虑$z$之趋向$z_0$.一旦固定了$N$,$S_N(z)$就是一个多项式,它的导数问题就化为有限多个$z^n$之求导,而这是我们已讲过的.因此,还是对上面那个$\varepsilon$,一定有$\delta>0$存在,使当$|z-z_0|<\delta$时第一项绝对值小于$\varepsilon/3$.综合起来,当$|z-z_0|<\delta$时

$$\left|\frac{f(z)-f(z_0)}{z-z_0}-f_1(z_0)\right|<\varepsilon.$$

即是说对适合$|z_0|<\rho$ 的任意 $z_0$，$f_1(z_0)=f'(z_0)$，从而在整个收敛圆$|z|<R$ 中，$f_1(z)=f'(z)$. 定理证毕.

读者可能会问，这里并没有出现一致收敛性. 确实如此，许多书上都把一致收敛性的发现归功于阿贝尔，其实事实并不如此. 阿贝尔知道，如果只使用简单收敛性是会出现毛病的. 可是阿贝尔说他自己是“够走运的”. 因为他“在分析里大抵总跟能表示成幂级数的函数打交道”. 一旦要与其他函数项级数打交道，只用简单收敛性就会出毛病了. 在 18 世纪，拉格朗日以为除幂级数以外再也没有其他级数了. 到 19 世纪，三角级数站了出来逼迫人们研究它. 阿贝尔的走运在于他扔下了傅里叶开辟的三角级数的研究，而找到了幂级数这个“安全地带”(阿贝尔的原话). 而一致收敛概念是赛德尔(1847)提出的，又经过魏尔斯特拉斯才为数学界普遍接受. 但是我们又可以说，这里确实有了一致收敛性，这就是看出了单用一个 $R$ 还不行，还要有一个 $\rho$ 使$|z|<\rho<R$. 如果没有 $\rho$，我们就不得不考虑整个收敛圆，而(3)在其上一般说来确非一致收敛的. 有了 $\rho$，我们实际上就是限制 $z$ 在一个紧集 $K$ 上. 这一点正是这里的证明的关键，也正是只有一致收敛才能保证的.

**注** 由这个定理立即可知，幂级数(3)可以逐项求导任意多次，而不改变收敛半径.

现在我们要给出一个极重要的概念.

**定义 1** 设 $f(z)$ 定义于一个开集 $D$ 上，而且在对 $D$ 中每一点都有连续导数，则称 $f(z)$ 在 $D$ 中解析.

于是定理 3 告诉我们，一个幂级数(3)在其收敛圆内表示一个解析函数. 令人吃惊的是，可以证明凡在开集 $D$ 中可求导的函数都可以展开为泰勒级数！但是，要能展开为泰勒级数至少要有一切阶导数，否则就不会有泰勒级数. 不但如此，还需要有许多别的结论：例如 $\overline{\lim\limits_{n\to\infty}}\sqrt[n]{|f^{(n)}(a)/n!|}<+\infty$，否则(1)不会有非 0 收敛半径. 如此等等. 看来，复域中的可求导性是一个内涵极为丰富的概念. 确实如此，我们在积分学一章中将会证明，只要 $f(z)$ 在开集 $D$ 中连续可导，则它在任一点 $a\in D$，均有收敛的泰勒级数(1)，而且只有通过积分才能把微分学的概念与幂级数展开式联系起来. 这确实是令人吃惊的事，它告诉我们复变量的解析函数理论是一个非常丰富，且在应用上也极重要的理论——我们通常称为“复分析”. 由于这个概念的重要性，我们多加一些说明.

在大多数书中，特别是关于复分析的著作中，解析性的定义只要求函数在 $D$ 中有导数而不要求有连续导数. 然后，在所有这些书中最后也都证明了，只要有导数，则导数必连续，这一点与实变量函数大异其趣. 但是证明这一点绝非易事，不仅这件事本身难证，而且用到解析性的一些根本定理都变得很难证了. 其实，在历史上，讲到解析性先是用我们这个定义的，到 20 世纪初才发现不要导数的连续性也行. 这当然包括了相当大的技巧，但最终并无新的收获，所以我们宁可用历史上老的定义.

然而在所有的关于复分析的书中都是说某个复变量函数 $f(z)$ 在一个开区域、一个开集、或某点的一个邻域(恒指开邻域)中解析，而没有说到在“某一点解析”、在“某一闭集上解析”，除非有专门声明的含意. 这与函数的连续性，可导性(即导数存在)完全不同，因为我们完全可以讨论函数的某一点的连续性，可导性等等. 这样一来，解析于 $a$ 点附近的解析函数就一定是指在某个小圆$|z-a|<r$ 中解析，而可以在 $a$ 点展开为幂级数(1)，它有一个正的收敛半径 $R$(与 $r$ 没有

直接关系)，并由此就有许多性质，也有研究它们的方法. 反之，即令我们规定了什么是 $f(z)$ 在某点的解析性，也看不出它有什么用. 回到我们上面的讨论，可见我们是定义了 $f(z)$ 在收敛圆 $D$: $|z-a|<R$ 中为解析. $f(z)$ 是否在收敛圆的圆周 $\partial D$: $|z-a|=R$ 上也解析呢? 当然这里所谓在 $\partial D$ 上某点处解析就理解为在该点的一个小邻域中解析，这是不可能的. 如果 $f(z)$ 在 $\partial D$ 之每一点上都按这个意义成为解析的，可以证明它一定在一个比 $D$ 更大的圆 $|z-a|<R_1$ $(R_1>R)$ 内解析. 这证明需要利用有限覆盖定理(即 Heine-Borel 定理). 这个定理将在第六章中详细讨论. 这就证明了收敛半径不再是 $R$，而会不小于 $R_1$，证明见第四章. 从这个意义上讲，解析函数 $f(z)$ 的收敛圆周上至少有一个"奇点"存在. 这里的奇点是指不能在这点(设为 $a$)处写出幂级数(1)来，或者有某一个 $f^{(k)}(a)$ 不存在，变为 $\infty$……或者尽管它们都存在，(1)的收敛半径却为 0. 可见奇点的概念其实与函数解析性密不可分. 上面我们说了，幂级数在收敛圆周上的性状极为复杂，但归结到底就在于其上有奇点. 但是这里的奇点却与临界点是两回事.

现在再回到幂级数(3)的分析运算问题，讨论它是否可以逐项积分. 这里就遇到了一个根本困难，什么是复变量函数的积分? 我们将在第四章中专门讨论它，并且看见其中牵涉了一些在实变量函数的情况下没有的问题. 现在只好把它放下. 但是还可以问另一个问题: 对(3)能否逐项求原函数? 原函数是与导数相逆的概念，既然能对 $f(z)$ 讲它的导函数 $f'(z)$，当然就可以讨论原函数. 而且利用前面讲过的知识可知(3)之一般项 $a_nz^n$ 的原函数就是 $\frac{a_n}{n+1}z^{n+1}+C$. 任意常数 $C$ 的出现添了一些麻烦: 各项分别添上了不同的 $C$，级数就会被弄成一团糟了. 于是不妨把问题变得更清楚一些: 能否逐项求在 $z=0$ 处为 0 的原函数? 而且为方便起见就用 $\int_0^z(\cdot)\mathrm{d}z$ 这个记号. 不过这个记号暂时还不代表作为和的极限的积分，而代表原函数. 读者不要以为我们在故弄玄虚: 原函数与积分不是一回事么? 否则又何必叫做不定积分呢? 恰好不是一回事. 前者是求导数的逆运算，后者是和的极限. 两个本不相同的概念后来被发现还有联系，这是一件大事，是牛顿—莱布尼茨的伟大功绩，因此才被称为微积分的基本定理. 但读者绝不要以为那些陈述起来很简单，证明起来又很容易的东西都是简单而不足道的. 这就是一个例子. 问题在于，如果我们限于讨论连续函数，则凡连续函数 $f(x)$ 必为可积，且 $\int_a^x f(t)\mathrm{d}t$，作为和的极限，只要取可变上限，是它的一个原函数. 反过来，若函数 $F(x)$ 是某连续函数 $f(x)$ 的原函数，则 $f(x)$ 必为可积，且 $F(x)$ 是它的不定积分(即有可变上限的和的极限)，即 $F(x)=\int_a^x f(t)\mathrm{d}t+C$. 弄清楚这个在连续函数范围中有效的关系是柯西的功绩——柯西基本上只研究了连续函数的积分. 一旦数学的发展使得不连续函数进入视野，使得人们不得不开始讨论不连续函数的积分时(傅里叶级数的研究是主要动因之一，而黎曼的贡献起了决定作用)，这个问题的复杂性就显露出来了. 我们在第四章中将详细地分析这个问题，而且它将导致积分学发展到勒贝格积分理论的阶段. 至于在复域中，应该说情况更奇怪，只限于连续函数是绝对不行的，而我们不得不研究解析函数: 若 $f(z)$ 是解析的，则它必为可积的，而且其(作为和的极限的)积分就是它的原函数. 反过来，解析函数 $F(z)$ 必是某函数的原函数: $F'(z)=f(z)$，且 $f(z)$ 必为(作为和的极限的)可积函数，而且 $F(z)$ 是它们的不定积分. $f(z)$ 还是解析的，一旦没有了解析性，以上所说全都没有了. 这些讨论也见第四章. 现在我们把这些很复杂的关系表述为一个很简单的定理.

**定理 4**　若幂级数(3)之收敛半径为 $R>0$，其和为 $f(z)$，则

$$\int_0^z f(z)\mathrm{d}z=\sum_{n=0}^{\infty}\int_0^z a_n z^n\mathrm{d}z=\sum_{n=0}^{\infty}\frac{a_n}{n+1}z^{n+1},\tag{13}$$

且其收敛半径不变，$\int_0^z(\cdot)\mathrm{d}z$ 表示在 $z=0$ 时为 0 的原函数.

**证**　收敛半径不变很容易证明，因为 $\sqrt[n]{\dfrac{1}{n+1}}=\dfrac{1}{\sqrt[n]{n+1}}\to 1$，其余证明与定理 3 几乎完全相同. 记(13)之和为 $F(z)$，则由定理 3

$$\frac{\mathrm{d}}{\mathrm{d}z}F(z)=\sum_{n=0}^{\infty}\frac{a_n}{n+1}\frac{\mathrm{d}}{\mathrm{d}z}(z^{n+1})=\sum_{n=0}^{\infty}a_n z^n=f(z),$$

而且易见 $F(0)=0$，所以 $F(z)$ 是 $f(z)$ 在 $z=0$ 时为 0 的原函数，即 $F(z)=\int_0^z f(z)\mathrm{d}z$. 定理证毕.

由定理 3 得到一个重要结论.

**定理 5**　幂级数(3)当收敛半径 $R>0$ 时必为其和在其中心 $z=0$ 处的泰勒级数.

**证**　若(3)之收敛半径为 0 则这个定理显然无意义. 当 $R>0$ 时，反复应用定理 3，有

$$f(0)=a_0,\qquad f'(0)=a_1,\cdots,\qquad f^{(k)}(0)=k!\ a_k,\cdots$$

因此定理得证.

这个定理之重要性在于它说明了，凡一个函数可以用幂级数(3)表示，则此级数必定就是其泰勒级数. 所以为了求 $f(z)$ 在 $z=0$ 处的泰勒展开式，我们不必去计算 $f(z)$ 之各阶导数，然后再计算泰勒级数的各个系数 $\dfrac{1}{k!}f^{(k)}(0)$，$k=0,1,\cdots$. 只要我们能用某种方法把 $f(z)$ 写成(3)，则(3)必是 $f(z)$ 以 0 为中心的泰勒级数. 不论用什么方法，只要能判断出(3)在某圆内收敛而在其外发散，此圆必为其收敛圆，而用不着去动用很难计算的阿达玛公式.

例如，利用定理 4 逐项求原函数有

$$\begin{aligned}\ln(1+z)&=\int_0^z\frac{\mathrm{d}z}{1+z}=\int_0^z\left[\sum_{n=0}^{\infty}(-1)^n z^n\right]\mathrm{d}z=\sum_{n=0}^{\infty}\frac{(-1)^n}{n+1}z^{n+1}\\&=\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}z^n,\qquad |z|<1.\end{aligned}\tag{14}$$

$$\arctan z=\int_0^z\frac{\mathrm{d}z}{1+z^2}=\int_0^z\left[\sum_{n=0}^{\infty}(-1)^n z^{2n}\right]\mathrm{d}z=\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}z^{2n+1},\quad |z|<1.\tag{15}$$

它们都恰好是泰勒级数，而收敛半径 $R=1$，事实上，当 $|z|>1$ 时，这两个级数的一般项都不趋于 0，因而都发散. 可是怎么知道 $\ln(1+z)$ 和 $\arctan z$ 恰好是 $\dfrac{1}{1+z}$，$\dfrac{1}{1+z^2}$ 的原函数呢？下一段中我们会回答这个问题.

**3. 解析延拓**　从(14)和(15)看见一个情况. 例如(14)中的函数 $\dfrac{1}{1+z}$，直接用定义易见它有连续的导数 $-\dfrac{1}{(1+z)^2}$，所以除了在 $z=-1$ 处以外，它是解析的. $z=-1$ 是它的奇点. $\dfrac{1}{1+z^2}$ 则

除了 $z=\pm\mathrm{i}$ 以外,处处都有连续导数.但是它们的以 $z=0$ 为中心的泰勒展开式却只在 $|z|<1$ 内适用.这就是说一个解析函数与这个解析函数的具体的表达式——泰勒级数只是表达式的一种——是有区别的.前者有自己的存在区域,后者作为它的局部的表达式只适用于一定的区域内.这里说到存在区域,就(14)与(15)两个例子中的被积函数来说,因为我们有一个非常简单的代数式子来表示它们,一眼就可看出,它们的存在区域分别是整个 $z$ 平面除去一点 $z=-1$ 或除去两点 $z=\pm\mathrm{i}$(还有无穷远点,将来读者也会看到是很容易处理的),但是 $\ln(1+z)$ 与 $\arctan z$,它们的存在区域又是什么呢?其实我们已经给出了它们的表达式为两个非常简单的积分(现在还只能说是原函数),可是怎样看出其存在区域是什么呢?读者们至此应该有一个感觉,如积分这样的式子,操作运算起来其实比某些"初等的"代数式更容易一些.可是要弄清由积分表示的解析函数的存在区域并非易事."从原则上说",我们可以应用更简单的泰勒级数.

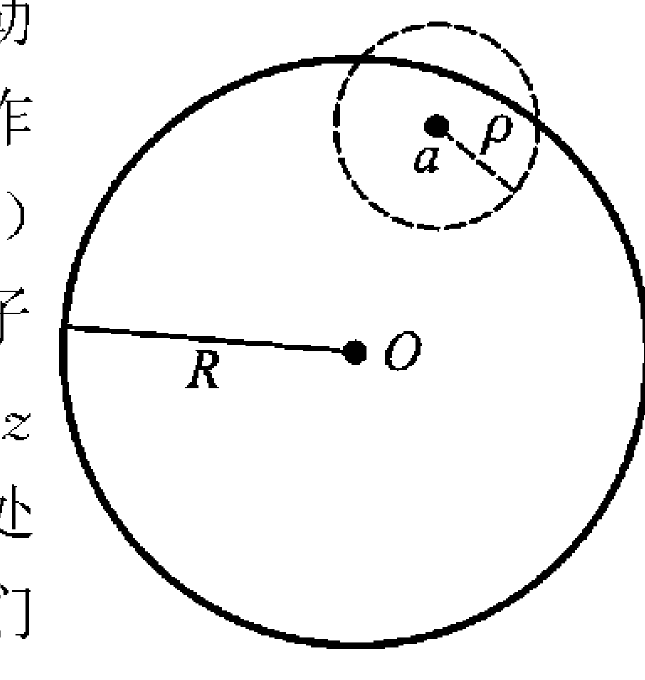


图 3-4-1

设有解析函数 $f(z)$,它在圆 $|z|<R$ 中可以表示为幂级数 $\sum_{n=0}^{\infty}a_nz^n$,$|z|<R$ 是它的收敛圆.在收敛圆中取一点 $a$,而且按定理 3,可以用此幂级数逐项求导来算出 $f^{(k)}(a)$,$k=0,1,2\cdots$,于是可以作出一个新的幂级数.

$$F(z)=\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(z-a)^n.$$

它的收敛圆形如 $|z-a|<\rho$,$\rho$ 的大小要使在圆周 $|z-a|=\rho$ 上,有 $f(z)$ 的奇点(这一点前面提出过,是需要证明的,但我们未证).此圆在图 3-4-1 上用虚线画出.在这两个圆的公共部分,$F(z)=f(z)$.但是可能有这样的情况,即 $|z-a|<\rho$ 有一部分位于 $|z|<R$ 之外如图.这时我们就说,$F(z)$ 是 $f(z)$ 由圆 $|z|<R$ 向 $|z-a|<\rho$ 的解析延拓.更一般地说,若有两个圆:$D_i$:$|z-a_i|<R_i$,$i=1,2$,以及两个各以它们为收敛圆的幂级数(3),分别记其和为 $f_1(z)$ 和 $f_2(z)$ 若 $D_1\cap D_2\neq\varnothing$,而且在 $D_1\cap D_2$ 上 $f_1(z)=f_2(z)$.令

$$F(z)=\begin{cases}f_1(z), & z\in D_1\\ f_2(z), & z\in D_2\end{cases}$$

我们称 $F(z)$ 是 $f_1$ 与 $f_2$ 向 $D_1\cup D_2$ 中的解析延拓,或者说 $f_1(z)(f_2(z))$ 是 $f_2(z)(f_1(z_1))$ 由 $D_2$ 向 $D_1$(由 $D_1$ 向 $D_2$)的解析延拓.我们这里采用了比较特殊的区域(圆盘)和解析函数比较特殊的表示(幂级数)来定义解析延拓,这样做的目的是为了避免一些几何上的困难.当然也可以用其它形状的区域或者解析函数的其他表示方式,但最后的结果都是一样的.不论怎么做,这里都会有两个问题.

第一个问题是唯一性问题.我们要证明,若 $f_1(z)$ 可以解析延拓到 $D_2$ 中成为 $\tilde{f}_2(z)$ 与 $\tilde{\tilde{f}}_2(z)$,或者说可以在 $D_1\cup D_2$ 中定义 $\tilde{F}(z)$,$\tilde{\tilde{F}}(z)$ 如上,而且 $\tilde{f}_2(z)=\tilde{\tilde{f}}_2(z)$ 于 $D_1$ 上,这时必有 $\tilde{f}_2=\tilde{\tilde{f}}_2$ 于 $D_2$ 上,或者 $\tilde{F}=\tilde{\tilde{F}}$ 于 $D_1\cup D_2$ 上.更广泛一点我们有

**定理 6(唯一性定理)**　若 $f(z)$ 在一个连通开集(即区域)$\Omega$ 中解析,而在 $\Omega$ 之一个非空开子集 $D$ 上 $f(z)=0$,则在

上

$$f(z)=\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(z-a)^n=0.$$

在此圆中取一点 $a_1$，又有 $f^{(n)}(a_1)/n!=0$，从而在某圆 $|z-a_1|<R_1$ 中又有 $f(z)=0$. 因为 $\Omega$ 是连通的，对任一点 $z_0\in\Omega$ 都可以找到一串（有限多个）如上的圆使 $z_0$ 落在某一圆中，在所有这些圆中，$f(z)=0$，因此 $f(z_0)=0$，而定理得证.

把这个定理应用于 $\tilde{F}-\tilde{\tilde{F}}$，即知 $\tilde{F}-\tilde{\tilde{F}}=0$ 于 $D$ 上，从而在 $D_2$ 上也有 $\tilde{F}-\tilde{\tilde{F}}=0$，亦即 $\tilde{f}_2(z)\equiv\tilde{\tilde{f}}_2(z)$.

上面的证明本质地依赖于 $\Omega$ 的连通性. 如果要把证明的细节都写出来，还相当繁冗，而与前面多次用过的连续拓展法是一样的. 我们不来重复这个证明，但是给出一个一望即知的“反例”说明连通性之不可少：令 $\Omega=D_1\cup D_2$，$D_1$，$D_2$ 是两个互相分离的区域，例如 $|z|<1$ 与 $|z-3|<1$. 令

$$f(z)=\begin{cases}0, & |z|<1,\\ 1, & |z-3|<1,\end{cases}$$

$f(z)$ 显然在 $\Omega$ 中解析，在它的一个非空开子集 $D_1$ 上为 0，但是并不恒为 0. 这个例子中的函数时常称为分片常值函数，时常用它来检验连通性的假设是否必不可少.

唯一性定理是解析函数最重要的性质之一而且十分有用，所以我们给出它的两个推广. 不但上述 $f(z)$ 不能在 $\Omega$ 之非空开子集 $D$ 上为 0，否则 $f\equiv 0$，而且如果 $f(z)$ 有一串零点 $\{z_k\}\to z_0$，且 $z_0$ 是 $\Omega$ 之内点，也会得到 $f(z)\equiv 0$. 换言之，若 $f(z)$ 在区域（即连通开集）$\Omega$ 中解析，且不恒为 0，则其零点在 $\Omega$ 中必是孤立的.

**推论 7**　若 $f(z)$ 在区域 $\Omega$ 中解析且 $f(z)\not\equiv 0$，若 $z_0\in\Omega$ 是 $f(z)$ 之零点，则必有 $z_0$ 的一个充分小邻域 $U$，使 $f(z)$ 在 $U$ 中除 $z_0$ 以外没有其他零点.

**证**　因为 $\Omega$ 是开集，$z_0\in\Omega$ 必为其内点，不妨设 $z_0=0$，因为 $f(z)\not\equiv 0$，故以 $f(z)$ 的 0 为心的泰勒级数必有第一个非 0 系数 $f^{(k)}(0)/k!=a_k\neq 0$，$k$ 为某正整数（注意 0 是 $f(z)$ 之零点）. 而 $f^{k-i}(0)/(k-i)!=a_{k-i}=0$，$i=1,2,\cdots,k$. 因此，这个泰勒级数可以写为

$$f(z)=a_kz^k[1+z\varphi(z)],$$

其中 $\varphi(z)$ 在 0 附近是解析的，因而在某个圆 $|z|\leqslant R$（含于收敛圆内）中连续而有界：$|\varphi(z)|\leqslant M$，取一个 $0<\rho<R$，使 $M\rho<\frac{1}{2}$，则在 $|z|\leqslant\rho$ 时

$$|f(z)|\geqslant|a_kz^k|[1-|z\varphi(z)|]\geqslant\frac{1}{2}|a_kz^k|,$$

所以推论得证.

**推论 8**　区域 $\Omega$ 中的不恒等于 0 的解析函数 $f(z)$ 只有有限阶零点.

**证**　用反证法. 设某一点 $a\in\Omega$ 是 $f(z)$ 之无限阶零点，则 $f^{(k)}(a)/k!=0$，$k=0,1,\cdots$. 所以 $f(z)$ 的以 $a$ 为中心的泰勒级数恒为 0，从而 $f(z)\equiv 0$ 于 $\Omega$ 中. 证毕.

解析延拓是一个我们时常用到而没有注意的概念. 例如在复域中我们用 (1

与 arctan $z$ 又该如何定义？以 $\ln(1+z)$ 为例，任取一点 $z_0$ 在 $|z|<1$ 中，并且按(14)计 $\ln(1+z_0)$，然后 $\frac{1}{1+z}$ 可以解析延拓到 $|z-z_0|<R$ 中，实际上

$$\frac{1}{1+z}=\frac{1}{1+z_0+z-z_0}=\frac{1}{1+z_0}\cdot\frac{1}{1+\zeta}=\frac{1}{1+z_0}\sum_{n=0}^{\infty}(-1)^n\zeta^n,$$

其中

$$\zeta=\frac{z-z_0}{1+z_0}.$$

此式当 $|\zeta|<1$ 即 $|z-z_0|<|1+z_0|$ 时有效. 因此在 $\{z;|z|<1\}\cap\{\zeta:|\zeta|<1\}$ 中

$$\ln(1+z)-\ln(1+z_0)=\int_{z_0}^{z}\frac{\mathrm{d}z}{1+z}=\int_0^{\zeta}\frac{\mathrm{d}\zeta}{1+\zeta}=\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\zeta^n,$$

所以

$$\ln(1+z)=\ln(1+z_0)+\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\left(\frac{z-z_0}{1+z_0}\right)^n.$$

但是右方的级数在 $\left|\frac{z-z_0}{1+z_0}\right|<1$ 亦即在 $|z-z_0|<|1+z_0|$ 中收敛，所以是 $|z-z_0|<|1+z$ 的解析函数. 而左方如果用(14)式表示，则只在 $|z|<1$ 中收敛. 所以(16)式是 $\ln(1+z)$ 由 $|$ 1 向 $|z-z_0|<|1+z_0|$ 中的解析延拓. 以上我们用了一些积分记号，但其实都是原函数的 我们用到的一些计算规则也都是原函数的计算规则，例如变量的变换等等. (16)这样的式子 告诉了我们 $\ln(1+z)$ 的解析延拓，但是使用起来并不方便，所以最好的办法仍然是看到，若 $=\ln(1+z)$ 表示解析延拓后的函数，由积分式得

$$\frac{\mathrm{d}w}{\mathrm{d}z}=\frac{1}{1+z},\qquad w\Big|_{z=z_0}=\ln(1+z_0).$$

就是说，可以用微分方程来定义一个解析函数的解析延拓. 在第二章中，我们都是用微分方 定义一些函数的，这样做的好处越来越明显了.

即得)，用二项级数，我们可以得到

$$\sqrt{1+z}=1+\frac{1}{2}z+\frac{1}{2!}\frac{1}{2}(\frac{1}{2}-1)z^2+\cdots.$$

它的收敛圆是 $|z|<1$，而在收敛圆周上有奇点 $z=-1$. 如果仿照第二章的作法，从 $D_1: |z|<1$ 依次作解析延拓回到 $D_N\equiv D_1$ 时将得到

$$\sqrt{1+z}=-\left[1+\frac{1}{2}z+\frac{1}{2!}\frac{1}{2}\left(\frac{1}{2}-1\right)z^2+\cdots\right]$$

(图 3-4-2). 在第二章中我们说，这个函数常被人们说成是多值函数，其实是黎曼曲面上的单值函数. 所以，一个如(3)那样的幂级数，经过一切可能的解析延拓，将得到一个定义于其黎曼曲面上的单值函数. 我们把这样得到的函数称为整体解析函数.

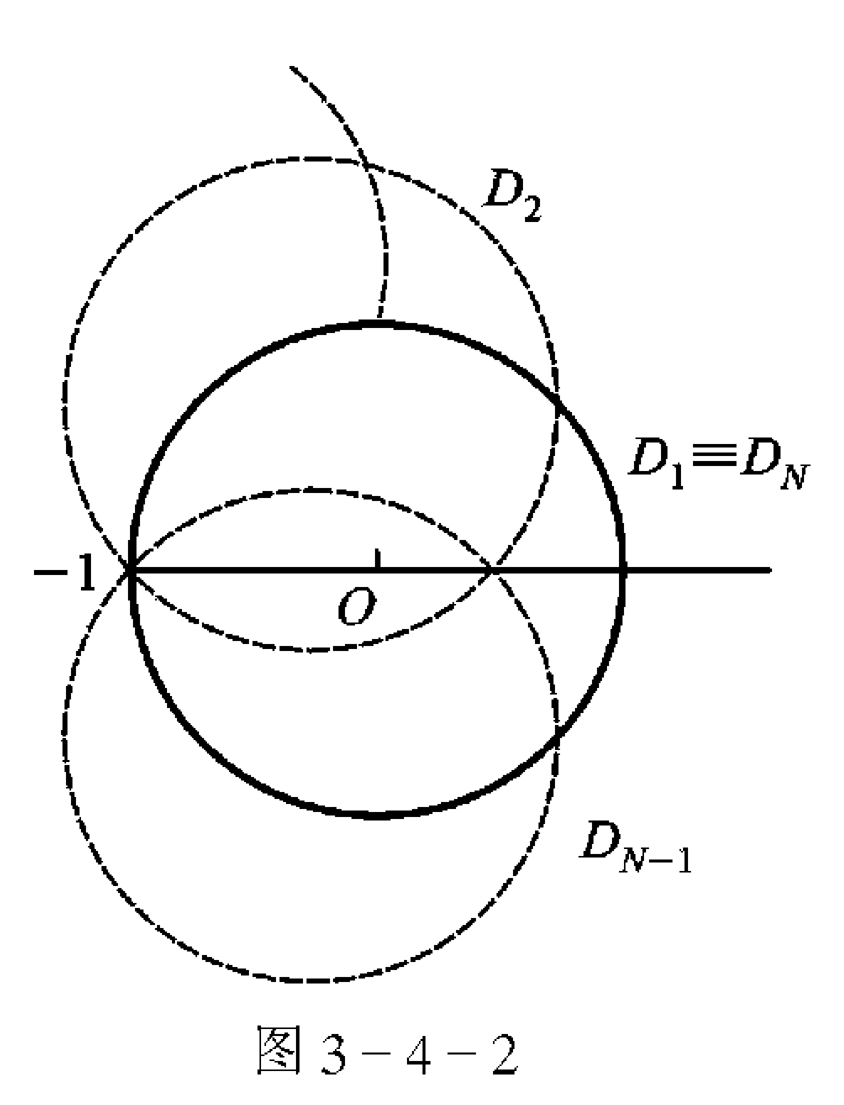


图 3-4-2

**定义 2** 从某一具有正收敛半径的幂级数(3)开始，经一切可能的解析延拓所得到的定义于其黎曼曲面上的函数称为一个整体的解析函数.

读者们可能会感到黎曼曲面是一个虽然有趣但是过于玄妙的东西. 不过我们要问一问读者，我们生活于其中的空间当真是那么简单吗？20 世纪下半叶的全部科学的发展都表明，空间是极为玄妙的，至今我们都说不清它的构造. 所以，如果有朝一日发现黎曼曲面有深刻的物理意义，或者有助于说明这个深刻的物理问题，我们不必感到吃惊.

**4. 柯西-黎曼方程、共形映射** 现在读者已经可以感到了，复变量函数的解析性是一个内涵非常丰富的概念. 复变量函数 $w=f(z)$ 是从 $z$ 平面的开集(我们只考虑开集) $\Omega$ 到 $w$ 平面的映射，可是 $z$ 平面也就是 $\mathbf{R}^2$ 平面 $z=x+\mathrm{i}y$，$w=u+\mathrm{i}v$ 平面也就是另一个 $\mathbf{R}^2$ 平面：$(u,v)$ 平面. 因此从几何上看 $w=f(z)$ 也是一个映射：$\Omega\subset\mathbf{R}^2\to\mathbf{R}^2$. 而 $f(z)$ 的解析性的要求就包括了 $u$ 与 $v$ 在 $\Omega$ 中有连续的偏导数 $u_x, u_y, v_x, v_y$. $f'(z)$ 的存在说明当 $z+h$ 以任何方式趋向 $z$ 时，$\frac{\Delta f}{\Delta z}$ 当 $\Delta z=h\to 0$ 时都有相同极限. 现在我们看 $h=h_1+\mathrm{i}h_2$，$h_1, h_2$ 为实数，以特定方式趋向 0 的后果. 一是 $h_1=0, h_2\to 0$，这时我们有

$$\lim_{h\to 0}\frac{\Delta f}{\Delta z}=\lim_{h\to 0}\left(\frac{u(x,y+h_2)-u(x,y)}{\mathrm{i}h_2}+\mathrm{i}\frac{v(x,y+h_2)-v(x,y)}{\mathrm{i}h_2}\right)$$
$$=\frac{1}{\mathrm{i}}\left(\frac{\partial u}{\partial y}+\mathrm{i}\frac{\partial v}{\partial y}\right).$$

同样，令 $h_2=0$，而 $h_1\to 0$，这时我们有

$$\lim_{h\to 0}\frac{\Delta f}{\Delta z}=\lim_{h_1\to 0}\left(\frac{u(x+h_1,y)-u(x,y)}{h_1}+\mathrm{i}\frac{v(x+h_1,y)-v(x,y)}{h_1}\right)$$
$$=\frac{\partial u}{\partial x}+\mathrm{i}\frac{\partial v}{\partial x}.$$

比较这两个式子即得：$f'(z)$ 存在而且连续意味着 $u, v\in C^1(\Omega)$ 而且

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y},\quad \frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}. \tag{18}$$

(18)式称为柯西-黎曼(Cauchy-Riemann，以下简记为 C-R)方程组，或称为 C-R 条件. 所以上

面我们证明了若 $f(z)$是 $\Omega$ 中的解析函数，则必有 $f\in C^1(\Omega)$而且其实部虚部适合 C－R 条件.读者自己也容易证明其逆：若 $f(z)=u+\mathrm{i}v$ 之实部虚部均为$C^1(\Omega)$函数，而且适合 C－R 方程组，则 $f(z)$必是 $\Omega$ 中的解析函数，因为容易证明$\lim\limits_{\Delta z\to 0}\dfrac{\Delta f}{\Delta z}$存在且连续，且$\dfrac{\partial u}{\partial x}+\mathrm{i}\dfrac{\partial v}{\partial x}$（或$-\mathrm{i}\dfrac{\partial u}{\partial y}+\dfrac{\partial v}{\partial y}$）即是这个极限，所以也可以用 C－R 条件来作为解析性定义的基础.

上面我们把 $w=f(z)$看成是 $\Omega\subset\mathbf{R}^2\to\mathbf{R}^2$ 的一个映射，但这个映射一般并不是微分同胚.因为它的雅可比行列式

$$J=\begin{vmatrix}u_x & u_y\\ v_x & v_y\end{vmatrix}=u_xv_y-u_yv_x=u_x^2+u_y^2=|f'(z)|^2 \tag{19}$$

可能在某些孤立点为 0.我们来看当 $J\neq 0$ 时这个映射的几何本质.在此映射下，线段$(z,z+h)$的像是一曲线，这个线段与实轴的交角是 $\arg h$.按微分学的基本思想，其像在忽略了一个与$|h|$比较为高阶的无穷小量之后，仍是直线，即过 $w=f(z)$点的切线（即图 3－4－3 之虚线）：

$$f(z+h)-f(z)=f'(z)h+o(h),\qquad\text{（曲线）}$$
$$w=f(z)+f'(z)h,\qquad\text{（切线）}$$

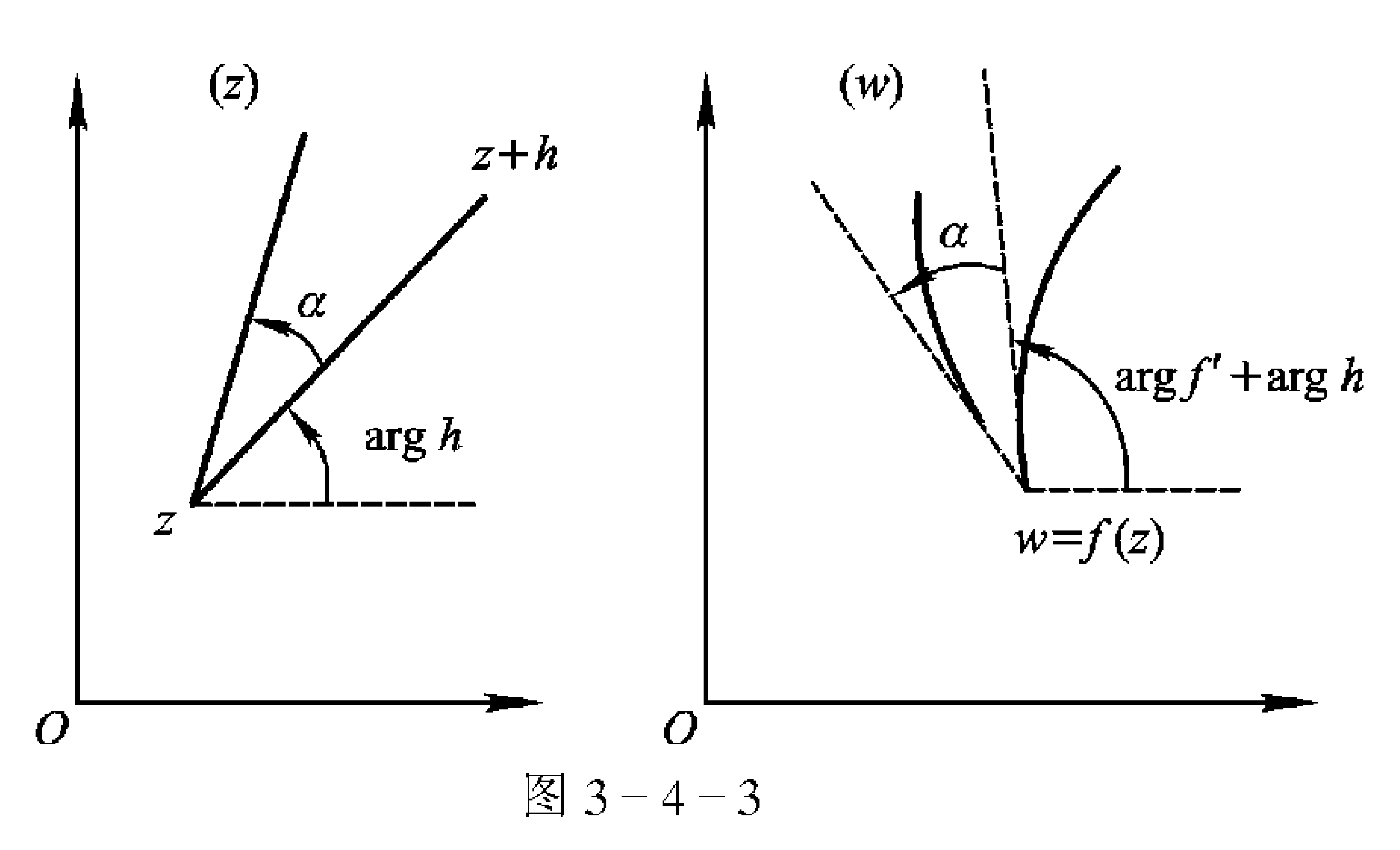



图 3－4－3

这条直线的倾角是

$$\arg f'(z)h=\arg f'(z)+\arg h.$$

所以原线段$(z,z+h)$之像，略去一个高阶无穷小量后，其方向是原直线段的方向再旋转一个角度 $\arg f'(z)$.在这里我们看到了 $f'(z)\neq 0$ 这个条件的作用.因为若 $f'(z)=0$ 则 $\arg f'(z)$没有意义.若过 $z$ 点作另一条直线段而与原直线段交角为 $\alpha$，则它在映射 $w=f(z)$下之像，过 $f(z)$点的切线也将旋转一个角度 $\arg f'(z)$而与上一直线段之像的交角仍为 $\alpha$.总之，由复变量的解析函数 $w=f$

$$\boldsymbol{E}=\mathbf{grad}\ u, \tag{20}$$

亦即

$$E_x=\frac{\partial u}{\partial x},\qquad E_y=\frac{\partial u}{\partial y}. \tag{21}$$

除了有电位 $u(x,y)$ 以外,还有一个重要的关系式如下:

如果在 $\Omega$ 内没有电荷,则由高斯定理知应有,

$$\operatorname{div}\boldsymbol{E}=\frac{\partial E_x}{\partial x}+\frac{\partial E_y}{\partial y}=0. \tag{22}$$

现在我们要介绍一个重要定理.众所周知,对适当光滑的函数 $\varphi$,$\frac{\partial^2\varphi}{\partial x\partial y}=\frac{\partial^2\varphi}{\partial y\partial x}$.所以,如果上面的 $E_x$ 与 $-E_y$ 分别是某一函数 $v$ 的偏导数,则上式自然成立.这里有一个重要定理(庞加莱引理)若 $\Psi_1,\Psi_2$ 适合 $\frac{\partial\Psi_2}{\partial x}=\frac{\partial\Psi_1}{\partial y}$,则它们一定是某个函数 $\Phi$ 的偏导数:$\Psi_1=\frac{\partial\Phi}{\partial x}$,$\Psi_2=\frac{\partial\Phi}{\partial y}$.这个结论的证明以及必须加上的条件将在第七章中详细讨论.将它用于 $E_x$ 与 $-E_y$ 即知,一定存在一个函数 $v(x,y)$ 使

$$E_x=\frac{\partial v}{\partial y},\ -E_y=\frac{\partial v}{\partial x}. \tag{23}$$

与(21)比较,即知,$u+\mathrm{i}v$ 适合 C-R 条件.因此,再附上适当的光滑性要求即知解析函数是描述平面静电场的适合的数学工具.其实,类似的情况还不只出现在静电学中,例如 18 世纪的欧拉(还有 18 世纪的达朗贝尔)在流体力学中也遇到过它,所以 C-R 方程也称为欧拉-达朗贝尔方程 特别是我们容易看到,由 C-R 方程可得

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0,\qquad \frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}=0.$$

即解析函数的实部虚部都适合拉普拉斯方程.而这是拉普拉斯在研究重力场的位势时得出的.

值得注意的是,本章§1 讲的几个数学物理中的偏微分方程的导出都比较简单,利用守恒律——甚至傅里叶犯了一个错误,把热当成一种流质来看待时,也还是在利用守恒律,然后再略去高阶无穷小量即得.现在却不同了.两次利用了位势的存在——后面用庞加莱引理得出 $v$ 时,虽然 $v$ 的物理意义并不清楚,但和电位的存在十分相似,这不是牛顿的微分学所能概括的.比之后来的麦克斯韦方程来说,静电场的研究还只是最简单的一步.我们在这本书中,还要追随着整个数学的发展,最后看到用数学来刻画电磁现象(包括光学)时,整个数学的发展已经远远地超越了牛顿.

还是回到本章主题,有了微积分,把它推广到复域是最直接也是最重要的一大发展.我们就看到了复变量的解析函数的内容多么丰富.等到我们看到了复域中的积分学又给我们带来什么惊喜的时候,对这一发展之大就更有体会了.

**5. $C^\infty$ 函数**　现在回到实变量 $x=(x_1,\cdots,x_m)$ 的 $C^\infty$ 函数的研究.与复变量的解析函数形成鲜明对比的是:后者(多个复变量情况也一样)只要知道其在某一点不论多么小的邻域中的值,则可通过解析延拓知道其在整个定义域上的值,不可以任加修饰.所以如果有一个复变量解析函数在某一点附近恒等于 0,则拓展后也只可能是恒为 0,此即唯一性定理.而 $C^\infty$ 函数,哪怕是毫不相干的,也可以通过稍加修饰拼在一起.这里重要的是应用一个在相当大区域中为 0 但又整体

上不恒为 0 的 $C^\infty$ 函数. 本节开始时讲过了一个. 但是以后用得最多的是

$$f(x)=\begin{cases}e^{\frac{1}{|x|^2-1}}, & |x|<1,\\ 0, & |x|\geqslant 1.\end{cases} \tag{24}$$

它在球面 $|x|=1$ 上也是 $C^\infty$ 的, 这是很容易证明的.

在往下讲之前, 先介绍一个概念和记号.

**定义 3** 设 $f(x)$ 定义于 $\Omega\subset\mathbf{R}^m$ 上, 我们称使得 $f(x)\neq 0$ 的点 $x$ 之集合的闭包为其支集, 并记为 $\operatorname{supp} f=\overline{\{x;\ f(x)\neq 0\}}$.

$\operatorname{supp} f$ 一定是相对于 $\Omega$ 的闭集. 要注意, 并不是 $f(x)$ 在 $\operatorname{supp} f$ 上不为 0, 例如定义在 $x\neq 0$ 处的 $f(x)=\sin\dfrac{1}{x}$, $f\left(\dfrac{1}{n\pi}\right)=0$, 但在 $\dfrac{1}{n\pi}$ 的任意小邻域中都有使 $\sin\dfrac{1}{x}\neq 0$ 的 $x$ 点, 所以 $\dfrac{1}{n\pi}$ 虽然使 $f(x)=0$ 却属于其支集, 而且是支集 $J=\{x;\ x\in\mathbf{R}, x\neq 0\}$ 的内点. 还要注意, 这个例子中的 $\Omega=\{x;\ x\in\mathbf{R}, x\neq 0\}$, $J$ 对于 $\mathbf{R}$ 并非闭集, 而是开集, 但是对于 $\Omega$ 却是一个闭集. 一个集合为开为闭需视相对于哪个集合而定. 这对我们理解何为开集, 何为闭集大有关系, 这就是子空间拓扑问题. 我们在第六章中会解释这个问题. 总之, $C^\infty$ 函数与解析函数是完全不同的对象.

我们下面用得最多的将是具有紧支集的 $C^\infty$ 函数. 所谓紧就是有界闭. $\Omega$ 的紧子集 $K$ 当 $\Omega$ 为开时必与 $\Omega$ 之边缘 $\partial\Omega$ 有一个正的距离. 这个正距离的存在关系极为重大. 从 §3 讲阿贝尔定理就可以看到这一点. 请读者务必十分注意. 具紧支集的 $C^\infty(\Omega)$ 函数之集合 (或称空间) 记为 $C_0^\infty(\Omega)$. 在不发生误会时就记为 $C_0^\infty$. (24) 的支集是单位球 $|x|\leqslant 1$, 但是我们可以作出支集几乎是任意集合的 $C_0^\infty$ 函数, 而且涉及数学分析中一个常用的技巧——磨光 (mollifying) 技术.

设紧集 $K\subset\mathbf{R}^m$, 而且其边界比较规则. 我们定义其特征函数 $\chi_K(x)$ 为

$$\chi_K(x)=\begin{cases}1, & x\in K;\\ 0 & x\notin K.\end{cases} \tag{25}$$

$K$ 之边界比较规则是为了保证 $\chi_K(x)$ 可积分. 规则到什么程度才能保证 $\chi_K(x)$ 可积, 下一章中关于多重积分的介绍中会讲到.

任取 $\delta\geqslant 0$, 我们定义

$$j_\delta(x-y)=\begin{cases}e^{\frac{1}{[|x-y|^2/\delta^2-1]}}, & |x-y|<\delta,\\ 0, & |x-y|\geqslant\delta.\end{cases} \tag{26}$$

于是

$$\int_{\mathbf{R}^m} j_\delta(x-y)\,\mathrm{d}y=\delta^m\int_{\mathbf{R}^m} e^{\frac{1}{|x|^2-1}}\,\mathrm{d}x=L\delta^m.$$

这里

$$L=\int_{\mathbf{R}^m} e^{\frac{1}{|x|^2-1}}\,\mathrm{d}x>0.$$

现在作积分

$$\Phi_\delta(x)=\frac{1}{L\delta^n}\int_{\mathbf{R}^m} j_\delta(x-y)\chi_K(y)\,\mathrm{d}y. \tag{27}$$

利用积分号下求导很容易看到 $\Phi_\delta(x)\in C^\infty$. 今证它有紧支集. 为此要注意，(27)式实际上只在 $K$ 上积分，所以(27)中的 $y$ 只在 $K$ 中. 注意 $x$，如果 $x$ 离 $K$ 的距离大于 $\delta$，即

$$x\notin K_\delta=\{x;\mathrm{dist}\,(x,K)\leqslant\delta\},$$

则(27)中的 $|x-y|\geqslant\delta$，从而 $j_\delta(x-y)=0$，而积分(27)为 0，所以 $\mathrm{supp}\,\Phi_\delta\subset K_\delta$，而 $\Phi_\delta\in C_0^\infty$.

$K_\delta$ 是一个比 $K$ 更大的紧集：它包含 $K$，而且我们记另一个比 $K$ 较小的紧集 $K^\delta$ 为含于 $K$ 内，且到 $\partial K$ 之距离不小于 $\delta$ 的集合为

$$K^\delta=\{x;\ x\in K,\mathrm{dist}\,(x,\partial K)\geqslant\delta\}.$$

设 $\delta$ 充分小时 $K^\delta\neq\varnothing$，则对 $x\in K^\delta$，以 $x$ 为心，$\delta$ 为半径之小球 $B_\delta(x)\subset K$，且

$$\Phi_\delta(x)=\frac{1}{L\delta^n}\int_{B_\delta(x)}j_\delta(x-y)\mathrm{d}y=1.$$

这个结果告诉我们当 $\delta\to0$ 时 $\Phi_\delta(x)$ 在某种意义下逼近 $\chi_K(x)$，因为除了在 $\partial K$ 附近的一个很"窄"的带形区域(即到 $\partial K$ 距离不到 $\delta$ 处)之外，$\Phi_\delta(x)$ 就是 $\chi_K(x)$. 这种逼近是很有用的. 实际上对任意连续函数 $f(x)$ 都可以用 $C_0^\infty$ 函数去逼近. 如果设 $f(x)$ 在 $\Omega$ 上连续，令

$$f_\delta(x)=\frac{1}{L\delta^m}\int_{\mathbf{R}^m}j_\delta(x-y)(\chi_\Omega f)(y)\mathrm{d}y \tag{28}$$

即可. 这里有一个小小的技术问题. 我们用 $\chi_\Omega f$ 代替 $f$，这样就不必在 $\Omega$ 上积分，而可以用全空间 $\mathbf{R}^m$ 上的积分了. 更重要的是，即令 $f$ 是 $\Omega$ 上的连续函数，它也不一定在 $\Omega$ 上可积. 因为 $f$ 在趋向 $\partial\Omega$ 时可以无界. 这就是我们在常用的微积分教本中讲的反常积分问题. 这时，我们通常说 $f$ 在区域 $\Omega$ 内局部可积，就是在含于 $\Omega$ 内的任一紧子集(即有界闭集)上可积. 但是即令如此，(28)与(27)仍有区别：(28)是在 $\Omega$ 上积分，$\Omega$ 不一定是紧集，其上的连续函数不一定可积，而(27)则是在 $K$ 上积分，$K$ 是我们假设为紧的. 不过我们要注意，这两个积分之被积函数还含有因子 $j_\delta(x-y)$，它只在球 $|x-y|\leqslant\delta$ 上不为 0，而积分(28)实际上是在球 $|x-y|\leqslant\delta$ 上积分的. 对于固定的 $x$，这确实是一个紧集. 于是我们有

**定理 9** 若 $f(x)$ 在开集 $\Omega$ 上连续，则对 $\Omega$ 内之任一紧子集 $K$，$C_0^\infty$ 函数 $f_\delta(x)$ 在 $K$ 上一致收敛于 $f(x)$.

**证** 因为 $\Omega$ 是开集，它的任一个紧子集 $K$ 离 $\partial\Omega$ 必有一个正距离 $\delta_0$，以下我们恒设 $\delta\leqslant\delta_0$. 以下恒设 $x\in K$.

因为 $\frac{1}{L\delta^n}\int j_\delta(x-y)\mathrm{d}y=1$，所以

$$f(x)=\frac{1}{L\delta^n}\int j_\delta(x-y)f(x)\mathrm{d}y.$$

现在 $K$ 离 $\partial\Omega$ 之距离 $\geqslant\delta_0$，而此积分中的 $y$ 又必须适合 $|x-y|\leqslant\delta$，否则 $j_\delta(x-y)=0$. 如果 $\delta<\frac{\delta_0}{2}$，则这样的 $y$ 必含于某个稍大于 $K$ 之紧集 $K_1$ 中，$K_1$ 离 $\partial\Omega$ 之距离 $\geqslant\frac{\delta_0}{2}$. 所以 $y\in\Omega$，$\chi_\Omega(y)=1$，$(\chi_\Omega f)(y)=f(y)$，而有

$$f(x)-f_\delta(x)=\frac{1}{L\delta^m}\int_{|x-y|\leqslant\delta}j_\delta(x-y)[f(x)-f(y)]\mathrm{d}y.$$

现在 $x$ 和 $y$ 全在紧集 $K_1$ 中，$f$ 在此紧集中必为一致连续的.所以 $\delta$ 充分小时 $|f(x)-f(y)|<\varepsilon$，$\varepsilon$ 是任意指定的正数.注意到 $j_\delta(x-y)\geqslant 0$，立即有

$$|f(x)-f_\delta(x)|\leqslant\frac{1}{L\delta^m}\int\limits_{|x-y|\leqslant\delta} j_\delta(x-y)|f(x)-f(y)|\mathrm{d}y<\varepsilon.$$

这就是说当 $\delta\to 0$ 时，$f_\delta(x)$在 $K$ 中一致收敛于 $f(x)$，证毕.

(28)式是一个积分算子：它以 $\frac{1}{L\delta^m}j_\delta(x-y)$ 为核，作用在 $\chi_\Omega f$ 上.这个算子称为磨光算子(mollifier)，$\frac{1}{L\delta^m}j_\delta(x-y)$ 称为磨光核.这个定理告诉我们，一个函数 $f$(乘上 $\chi_\Omega$ 是为了使它处理起来更方便)经过磨光就能得到一个 $C_0^\infty$ 函数族 $\{f_\delta\}$ 去逼近它.当 $f$ 是连续函数时，则可以使 $\{f_\delta\}$ 一致收敛于它.当 $f$ 是其他函数类之元时，则可以在相应的意义下使 $\{f_\delta\}$ 收敛于它.这是现代数学中为了克服函数不光滑性常用的方法.它告诉我们，$C_0^\infty$ 函数是很多很多的.

$C_0^\infty$ 函数另一个重要的用途是利用它来构造单位分解.这是把整体性问题分解为局部性问题，或把局部性结果拼接成整体性结果的工具，它的基础是 $\mathbf{R}^m$ 中紧集(即有界闭集)的一个重要性质.

**定理 10(海涅－博雷尔(Heine-Borel)定理或有限覆盖定理)** 设 $K\subset\mathbf{R}^m$ 为一紧集，$\{U_\lambda\}$ 是一族开集且覆盖 $K$：$\bigcup\limits_\lambda U_\lambda\supset K$，则从 $\{U_\lambda\}$ 中必可找出有限多个：$U_1,\cdots,U_N$ 仍可覆盖 $K$：$\bigcup\limits_{i=1}^N U_i\supset K$.

这个定理的证明见第六章.$\{U_\lambda\}$称为 $K$ 的一个开覆盖.这里参数 $\lambda$ 可以是有限多个，可数无穷多个，不可数无穷多个，总之是要覆盖 $K$.本章中我们已多次讲到紧集的概念，紧性是整个数学中最基本的概念之一.在通常的微积分教本中时常用有界闭集的说法来代替紧集，这是不恰当的.在第六章我们要详细讨论这个概念，到那时才能明白，在什么情况下才能用有界闭这个概念来代替紧性，也才能明白定理 10 为什么是一个需要证明的结果，它与微积分学中其它的基本概念又有什么关系和什么区别.又附带说一句，Borel 常译为波莱尔.

在承认了定理 10 之后，我们要在 $\mathbf{R}^m$ 中构造所谓单位分解.下面的记号与名词仍与定理 10 相同.

**定理 11(单位分解的存在定理)** 对于上述的 $K$，开覆盖 $\{U_\lambda\}$ 以及有限的子覆盖 $U_1,\cdots,U_N$，一定存在函数 $\Phi_j(x)\in C_0^\infty(\mathbf{R}^m)$，$j=1,\cdots,N$，适合以下要求

(1) 每个 $\Phi_j$ 的支集必在某一 $U_i$ 中 ($i$ 与 $j$ 不一定相同)；

(2) $0\leqslant\Phi_j(x)\leqslant 1$，　　$1\leqslant j\leqslant N$；

(3) $\sum\limits_{j=1}^N \Phi_j(x)=1$， (29)

$\{\Phi_j(x)\}$称为 $K$ 的从属于覆盖$\{U_i\}$的 $C_0^\infty$ 单位分解.

**证** 任一点 $x\in K$ 必含于某个 $U_i$ 中，从而也必含于一个开的小球 $K_j$ 中，这里 $K_j\subset U_i$，所以$\{K_j\}$也成了 $K$ 的一个开覆盖而必可以从其中选出有限多个：$K_1,\cdots,K_k$，使 $K\subset\bigcup\limits_{j=1}^k K_j$，而每

个 $K_j$ 含于一个 $U_i$ 中,这里 $i$ 与 $j$ 不一定相同,而且同一个 $K_j$ 可以含于好几个 $U_i$ 内,因而$\{K_j\}$之个数与$\{U_i\}$之个数也不一定相同.必要时稍为缩小 $K_j$ 可以设$K_j \subset U_i$,而 $\mathrm{dist}\,(K_j, \partial U_i) = \delta_j > 0$,于是用上面的方法对 $\chi_{K_j}(x)$作磨光,即用(27)式作 $\varphi_{\delta_j}(x) \in C_0^\infty$,很明显 $\chi_{K_j}(x) = 1$ 于 $K_j$ 中,而且必要时稍微缩小 $K_j$,当 $x \in K_j$ 时有

$$0 \leqslant \varphi_{\delta_j}(x) = \frac{1}{L\delta_j^m} \int j_{\delta_j}(x-y) \chi_{K_j}(y) \mathrm{d}y = \frac{1}{L\delta_j^m} \int j_{\delta_j}(x-y) \mathrm{d}y = 1.$$

因为 $\bigcup_j K_j \supset K$,故对任意 $x \in K$,必至少有一项 $\varphi_{\delta_j}(x) = 1$,从而

$$\sum_j \varphi_{\delta_j}(x) > 0.$$

令

$$\Phi_j(x) = \varphi_{\delta_j}(x) \Big/ \sum_j \varphi_{\delta_j}(x),$$

易见$\{\Phi_j(x)\}$即是所求的单位分解.

以上我们只对紧集的开覆盖证明了单位分解的存在.实际上即令 $K$ 不是紧集,单位分解也是有的,但其概念要稍作微改.详细的讨论见第六章和专门的参考书籍.

以上我们讲的是 $C_0^\infty$ 单位分解.也可以构造 $C_0^k$ 单位分解,但是不可能作出解析的单位分解,因为解析函数除非恒为 0 是不会有紧支集的.这里也表现了解析函数与 $C^\infty$ 函数的区别.

## §5 反函数定理和隐函数定理

**1. 重温线性代数** 前两节中我们讨论了映射 $f: \Omega \subset \mathbf{R}^m \to \mathbf{R}^n$ 所生成的无穷小量$f(x+h) - f(x)$按 $h$ 之幂分解的问题,并由此引出了解析函数与 $C^\infty$ 函数的区别的详细讨论.现在再回到 $f(x+h) - f(x)$之线性部分即一阶微分的讨论.我们在前面一再强调线性部分的研究可以提供有关这个映射的局部的(即在 $x$ 点附近,以下我们限于在原点附近)性态的信息.现在再回到这个主题并讨论一个极重要的问题即如何求解

$$f(x) = y. \tag{1}$$

这个问题把我们引导到一个更广泛的问题,即讨论 $y = f(x)$的逆映射 $f^{-1}$.更准确一些说是讨论 $y \in \mathbf{R}^n$ 是否在映射 $f$ 下有原像.因为只有在这个原像是存在而且唯一的条件下,才能说有逆映射 $f^{-1}$.

现在在 $x = 0$ 附近把 $f(x)$线性化,即令

$$f(x) = f(0) + Ah + o(h)$$

并用 $f(0) + Ah$ 取代$f(x)$.这里的 $A$ 是由$\Omega$ 的切空间 $T\Omega = \mathbf{R}^m$ 到 $\mathbf{R}^n$ 的切空间(仍为 $\mathbf{R}^n$)的线性映射.如果用坐标表示,则有

$$A = \frac{\partial(f_1, \cdots, f_n)}{\partial(x_1, \cdots, x_m)}(0).$$

以下为方便起见,恒设 $f(0) = 0$.于是映射 $y = f(x)$将被线性映射

$$k = Ah, k \in \mathbf{R}^n, h \in \mathbf{R}^m \tag{2}$$

取代.研究(2)是线性代数学的主要问题之一.因为我们在前几节中一再强调用一种与坐标无关的方式来研究映射(1),所以现在我们也希望用一种与坐标无关的方式来研究(2).下面我们限于讨论 $A$ 的核空间与像空间.

研究方程(2)对于某一个 $k$ 是否有解 $h$ 存在,也就是研究 $k$ 是否在 $A$ 的像空间中,即问是否有 $k\in imA$:

$$imA=\{k\in \mathbf{R}^n, \exists h\in \mathbf{R}^m \text{ 使 } k=Ah\}. \tag{3}$$

$imA$ 是一个线性空间,它并不一定就是整个靶空间 $\mathbf{R}^n$ 而可能只是它的一个真子空间.$imA$ 与 $\mathbf{R}^n$ 之差别可以用它们的维数来刻画.$\dim\mathbf{R}^n=n$ 是固定的,而 $\dim imA$ 称为 $A$ 之秩(rank),记为 $\text{rank } A$.秩有一个重要作用,即用它来判断(2)的可解性最方便.事实上,(2)对一切 $k\in\mathbf{R}^n$ 均有解,即(2)为可解(solvable)之充分必要条件当然就是靶空间与像空间相同,亦即它们有相同维数,所以 $n=\text{rank } A$ 是(2)对一切 $k$ 可解的充分必要条件.这时 $A$ 把源空间 $\mathbf{R}^m$ 映到整个靶空间上,所以 $A$ 称为满射(surjection).更简明的说法就说这是 $A$ 是一个 onto.因为在一般情况下 $A$ 把源空间映到靶空间内,所以是一个 into.另一个重要的子空间是源空间中一切被 $A$ 映到靶空间中的 0 的向量所成,它称为 $A$ 之核(kernel)空间或简单就是核.记作 $\ker A=\{h\in\mathbf{R}^m, Ah=0\}$.如果 $\dim\ker A=0$ 即是说 $\ker A$ 中只有一个零向量 0,则方程(2)若有解也只能是唯一的.因为若 $h_1, h_2$ 都是(2)的解,当有 $A(h_1-h_2)=k-k=0$.所以 $h_1=h_2$,而(2)之解是唯一的.反过来,若(2)之解为唯一的则 $\ker A=\{0\}$.实际上,若有 $0\neq h_0\in\ker A$,则对(2)之每一个解 $h$,$h+Ch_0$($C$ 为任意常数)也是(2)之解,这与(2)之解的唯一性矛盾.总之,(2)有唯一解当且仅当 $\ker A=\{0\}$.这时我们称 $A$ 为单射(injection).如果 $A$ 既是单射又是满射,就称为一一映射(bijection),也称双射.一一映射就是 $A$ 实现了源空间 onto 靶空间的同构.这时 $A^{-1}$ 才存在,而且也是同构.

以上我们是把 $\text{rank } A$ 与靶空间 $\mathbf{R}^n$(以下记为 $F$)的维数联系起来看的.由秩之定义,自然有 $\text{rank } A\leqslant\dim F$.但还有一个重要公式,把 $\text{rank } A$ 与 dim(源空间)联系起来.源空间现在就是 $\mathbf{R}^m$,我们也不妨记为 $E$.我们知道,由于 $\ker A$ 中可能有非零元,所以 $A: E\to imA$ 虽然是 onto,但不一定是一对一的,即非单射.但是通常的线性代数教本讲到商空间时,一定会讲到一个结论:$A$ 在商空间 $E/\ker A$ 与 $imA$ 之间诱导出一个同构(仍记为 $A$).所以它们的维数相同.但是 $\dim(E/\ker A)=\dim E-\dim\ker A$,所以有

$$\dim E-\dim\ker A=\dim imA. \tag{4}$$

由于下面常用这个结果,所以我们把它写成一个

**引理 1** 设 $A:\mathbf{R}^m\to\mathbf{R}^n$ 是一个线性映射,则

$$\dim\ker A+\text{rank } A=m. \tag{5}$$

**证** 现在源空间 $E=\mathbf{R}^m$,故 $\dim E=m$,$\dim imA$ 按定义即 $A$ 之秩 $\text{rank } A$,代入(4)即得.

由此立即有 $\text{rank } A\leqslant m$,但由定义 $\text{rank } A\leqslant n$,所以

$$\text{rank } A\leqslant\min(m,n). \tag{6}$$

当我们对线性映射 $A$ 之核空间与像空间作了如上的讨论以后,就可以再来看方程(2)的求解问题.最好的情况显然是 $A:\mathbf{R}^m\to\mathbf{R}^n$ 为同构的情况.这时 $A$ 首先是一个满射,所以 $\text{rank } A=n$.另一方面又是单射,所以 $\dim\ker A=0$,而由(5)式 $\text{rank } A=m$,所以 $A$ 为同构必导致

$$m = n = \text{rank } A. \tag{7}$$

反过来由(7)，一方面 rank $A = \dim imA = n$，故 $A$ 为满射，另一方面 $\dim\ker A = m - \text{rank } A = 0$ 故 $A$ 为单射，总之 $A$ 为同构. 这样可以看到(7)是 $A$ 为同构的充分必要条件.

由(7)看到，若 $n \neq m$，则 $A$ 不可能是同构. 下面就分别来讨论 $m < n$ 与 $m > n$ 的情况. 但不论如何，(7)是一种秩已达到最大可能的允许值的情况. 所以下面我们在每个情况下都只讨论 rank $A$ 达到最大可能的允许值的情况. 由(6)可知，这就是 $m < n$ 而 rank $A = m$，以及 $m > n$ 而 rank $A = n$.

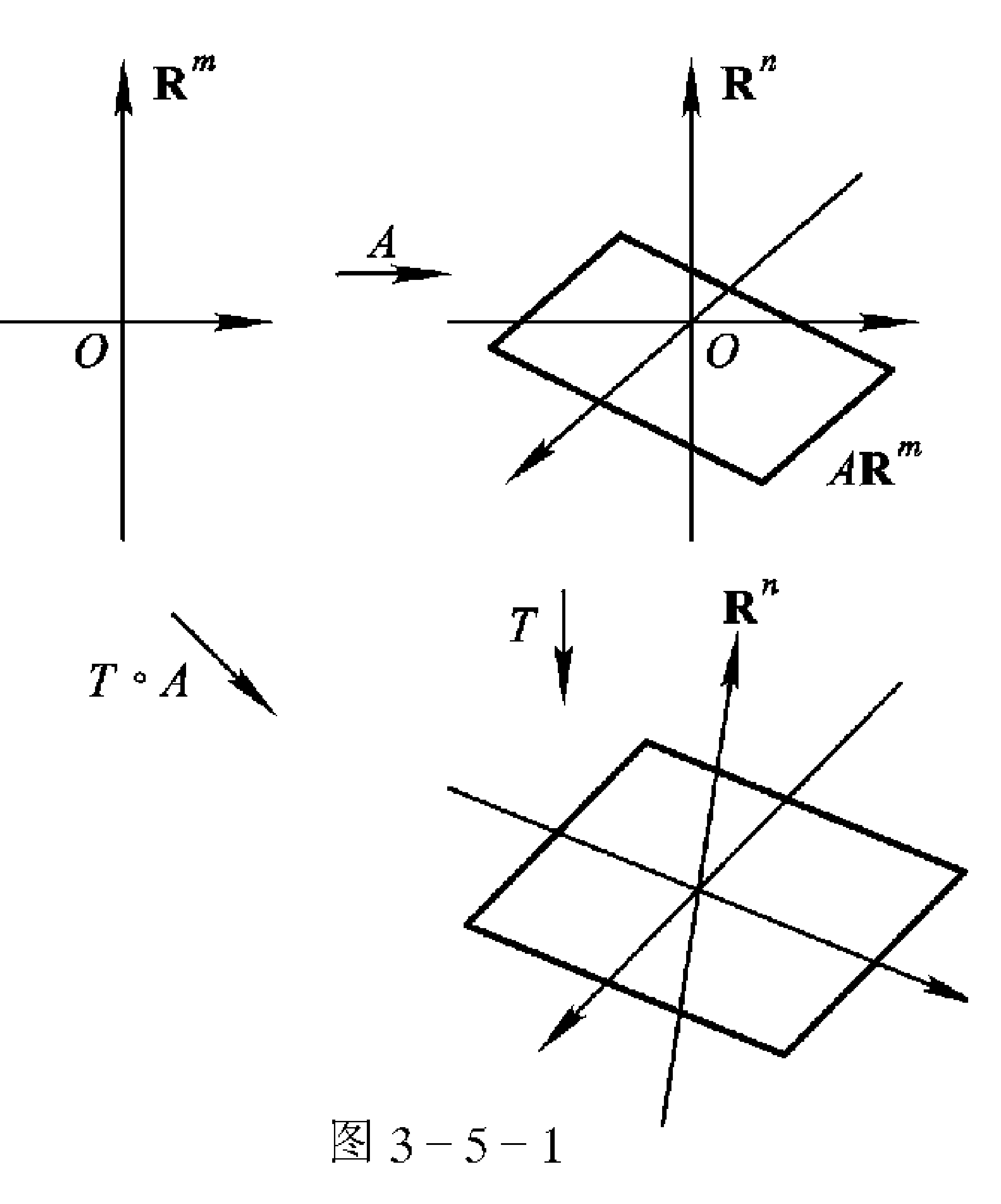


图 3-5-1

先看 $m < n$，rank $A = m$. 这时，$A$ 把源空间 $\mathbf{R}^m$ 映为靶空间的 $m$ 维子空间，而且由(5)知 $\dim\ker A = 0$，所以 $A$ 是一个单射. 图 3-5-1 上一行画出了这个映射. 然而我们可以在靶空间中作了一个坐标变换 $T$ 使得 $A\mathbf{R}^m$ 成为一个"坐标平面". 图 3-5-1 右侧一列画出了 $T$，这样我们就可以想像到，在这样的坐标系下，$A$ 应该有特别简单的表示. 下面我们引用坐标表示，并且设在源空间与靶空间中都已选好了坐标系——即基底，从而 $h$ 和 $k$ 都表示为列(竖)向量：

$$h = {}^t(h_1, \cdots, h_m), \qquad \text{(上角的 } t \text{ 表示转置.)}$$
$$k = {}^t(k_1, \cdots, k_n),$$

而 $A$ 表示为

$$A = (a_{ij}),\ i = 1, 2, \cdots, n,\ j = 1, 2, \cdots, m.$$

如果在靶空间 $\mathbf{R}^n$ 中作一个坐标变换，而把 $k$ 变成 $l$

$$l = {}^t(l_1, \cdots, l_n).$$

应当有

$$l = Tk = (T\circ A)h,$$

$T = (t_{ij})$是一个 $n \times n$ 非异矩阵($n$ 阶非异方阵所成的空间在数学中极为重要. 这个空间按习惯记为 $GL(n, \mathbf{R})$. $n$ 表示方阵阶数，$\mathbf{R}$ 表示其元为实数). 原来的 $\mathbf{R}^m = \{(h_1, \cdots, h_m)\}$，而现在的 $(T\circ A)\mathbf{R}^m = \{(l_1, \cdots, l_m; 0, \cdots, 0)\}$. 因此经过坐标的选择，映射 $A$ 就成了

$$l_i = h_i, \qquad i = 1, 2, \cdots, m,$$
$$l_{m+j} = 0, \quad j = 1, \cdots, n - m. \tag{8}$$

如果用矩阵来表示(8)，它就是

$$\begin{pmatrix} I \\ O \end{pmatrix}. \tag{9}$$

这是一个 $n \times m$ 矩阵，而它的子矩阵 $I$ 则是 $m \times m$ 的单位矩阵. 从几何上看，这就是把源空间"浸入"或"嵌入"在靶空间中. 但是在数学中，浸入(immersion)与嵌入(embedding)这两个名词，虽然是描述的这种映射，但都有确切的定义，而且二者又有比较细微但十分重要的区别. 它们属

于微分拓扑学研究的范围,本书只能在第七章中稍为涉及.所以现在不用这些名词,而只说它是"单射"线性映射,所以(8)就是单射线性映射的标准形式.

其次看 $m>n$,而 $\operatorname{rank} A=n$ 的情况.这是由一个高维空间向低维空间的映射.图 3-5-2 上画的是 3 维空间向 2 维空间的映射.为了简化它,我们先把 $\mathbf{R}^m$ 中的坐标系作一个倾斜成为下面用虚线画出的坐标系,使得 $\mathbf{R}^n$(现在是图右半部的平面)恰好成为虚线坐标系的"坐标平面".我们把倾斜映射记为 $T^{-1}$,于是方程(2)现在成为

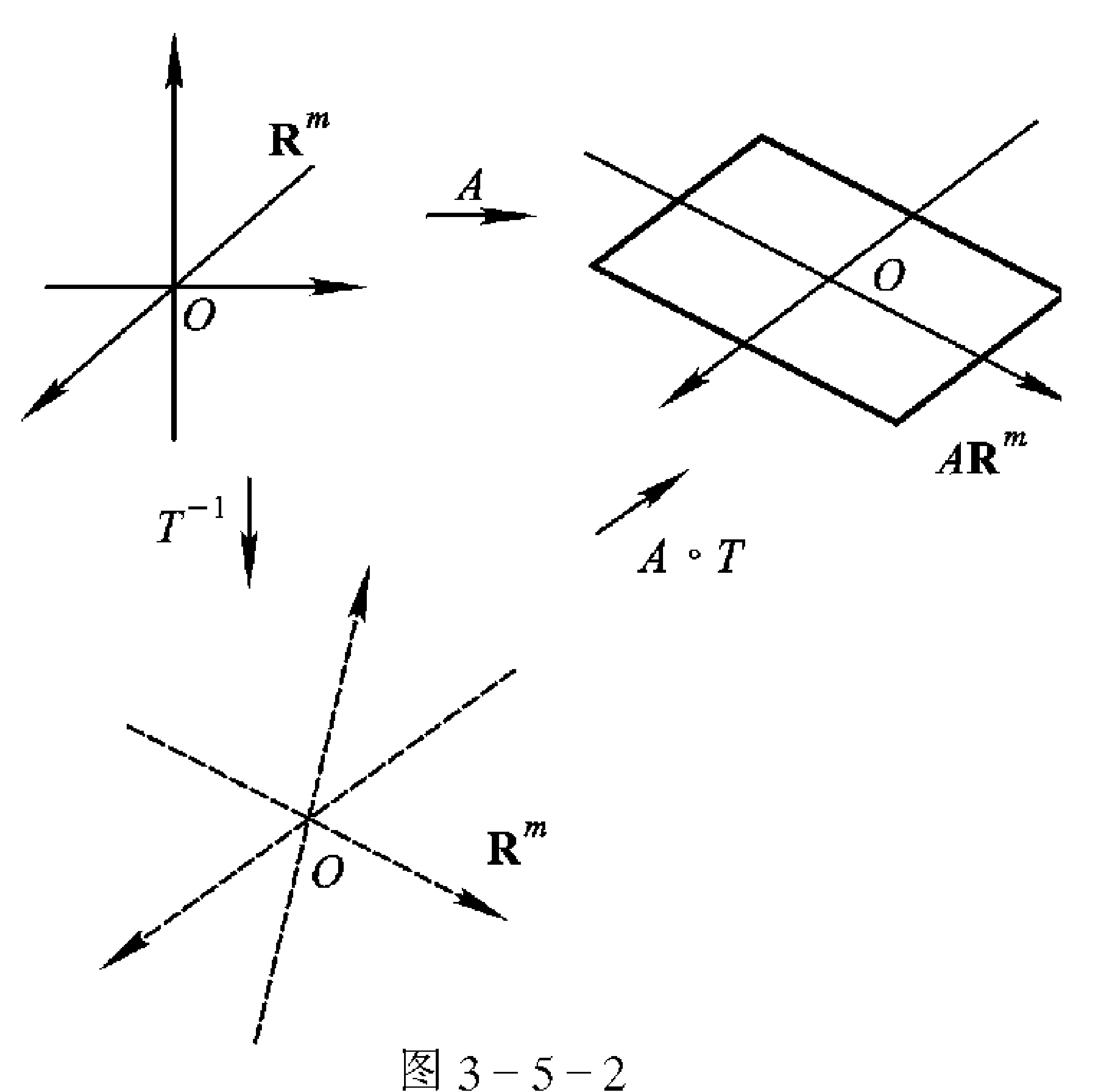



图 3-5-2

$$k=Ah=(A\circ T)\circ(T^{-1}h)=(A\circ T)l. \quad (10)$$

如果记 $l={}^t(l_1,\cdots,l_{m-n};l_{m-n+1},\cdots,l_m)$,它一共有 $m$ 个分量而 $\mathbf{R}^n$ 上之向量一共只有 $n$ 个分量.从图上看,恰好是 $l$ 的后 $n$ 个分量.因此,$A\circ T$ 如果用坐标来写就是

$$h_i=l_{m-n+i} \qquad i=1,2,\cdots,n. \quad (11)$$

如果用矩阵来表示它,这个矩阵就是 $n\times m$ 矩阵

$$(O \quad I),$$

其中的 $I$ 是一个 $n\times n$ 的单位矩阵.那么要问什么样的向量被 $A$ 映为零向量?即问 $\ker A=A^{-1}(0)$是什么样的空间?由(11)易见一个向量 $l=(l_1,\cdots,l_{m-n},l_{m-n+1},\cdots,l_m)\in\ker A$,当且仅当其后面的 $m$ 个分量为 0.即是说 $\ker A$ 是由形如

$$(l_1,\cdots,l_{m-n},0,\cdots,0)$$

的向量所张成的 $m-n$ 维子空间.

用直观的语言来说,这就是:从倾斜的坐标系来看,就好像是把斜着向上的 $m-n$ 坐标被"压"掉了,$\ker A$ 是这 $m-n$ 个坐标向量张成的,而 $A$ 把像空间 $\mathbf{R}^n$"罩"起来了.所以它称为"外罩"(submersion,或译"浸没",指靶空间沉没在 $A$ 的像中去了,但"浸没"一词反而用得多一些.第七章都是用的浸没).也和上面一样,由于外罩是一个有特定含意的数学名词.我们就不去使用它,而说这是一个"投影"(projection)线性映射,准确些说,是"沿"着 $\ker A$ 向 $imA$ 投影.

以上所说的是 $m>n$ 与 $m<n$ 的情况,但我们也可以把 $m=n$ 算起去,它既是单射,又是投影.因此它相应的矩阵就是 $I$.

总之,$\operatorname{rank} A$ 取可能最大值的线性映射分成三种:

1. $m=n$,这时 $A$ 就是同构,而在适当坐标系中其矩阵即是 $I$.

2. $m<n$,这时 $A$ 是单射映射(8),而在适当坐标系中,它的矩阵是 $\begin{pmatrix} I \\ O \end{pmatrix}$.

3. $m>n$,这时 $A$ 是投影映射(11),而在适当坐标系中它的矩阵是$(O \quad I)$.

本节的主要内容就是证明:可微映射在秩取最大可能的值时,局部地正是这三种情况.

**2. 反函数定理** 现在我们要证明下面的基本定理

**定理 2(反函数存在定理)** 设 $\Omega\subset\mathbf{R}^n$ 是一开集,$f:\Omega\to\mathbf{R}^n$ 是一个光滑映射,$0\in\Omega$ 且 $f(0)=0$.若$(\mathrm{d}f)(0)$是非退化的,则 $f$ 必是 $0\in\Omega$ 的某邻域到 $f(0)\in\mathbf{R}^n$ 的某邻域中的微分同胚.

在给出证明之前,先对定理的含意作一些注解.首先要注意,这个定理是局部的,它所给出的微分同胚也只是局部微分同胚.微分同胚的概念已见于本章§2定义3.引理4后的注1就提出了一个映射为微分同胚的充分必要条件问题.从那里就已看到,若用局部坐标来表示,则 $(\mathrm{d}f)(0)=\frac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_n)}(0)$,$(\mathrm{d}f)(0)$之非退化性就表示为雅可比矩阵非退化.亦即其行列式不为0.但那里即已指出,这并非充分条件,而只局部地才是充分必要条件.本定理就给出这个论断的证明.如果把本定理中的映射 $f$ 用局部坐标写为

$$y_i=f_i(x_1,\cdots,x_n),0=f_i(0,\cdots,0),i=1,2,\cdots,n, \tag{12}$$

局部微分同胚就表明,$x_i$ 也是 $(y_1,\cdots,y_n)$ 的光滑函数

$$x_i=\varphi_i(y_1,\cdots,y_n),0=\varphi_i(0,\cdots,0),i=1,2,\cdots,n, \tag{13}$$

而以(13)代入(12)或以(12)代入(13)均有

$$y_i\equiv f_i[\varphi_1(y),\cdots,\varphi_n(y)],i=1,2,\cdots,n,$$

或

$$x_i=\varphi_i[f_1(x),\cdots,f_n(x)],i=1,2,\cdots n.$$

这就是说,(12)与(13)互为反函数.因此这个定理是反函数存在定理.

以上我们一再说坐标.现在也应该把坐标的定义说明白.对于线性空间 $\mathbf{R}^n$,所谓坐标是指有了一个基底 $\tilde{e}_1,\cdots,\tilde{e}_n$ 以后,任一点亦即任一向量 $\tilde{x}$ 均可写为 $\tilde{x}=x_1\tilde{e}_1+\cdots+x_n\tilde{e}_n$. $(x_1,\cdots,x_n)$ 即此向量 $\tilde{x}$(或记为此点 $x$)的坐标.现在我们讨论的 $\Omega$ 虽然是 $\mathbf{R}^n$ 的一个开子集,却不是线性空间,这时所谓坐标即指 $\Omega$ 中任一点均有一个邻域与 $\mathbf{R}^n$ 同胚(即双方一对一且双方连续).设 $P$ 对应于 $\mathbf{R}^n$ 中一点 $x=(x_1,\cdots,x_n)$,就说 $x$ 或 $(x_1,\cdots,x_n)$ 是 $P$ 的局部坐标.因为这种同胚有许许多多,$P$ 就有许多不同的局部坐标例如 $x=(x_1,\cdots,x_n)$,$y=(y_1,\cdots y_n)$ 而 $y=y(x)$ 或 $x=x(y)$.我们限于考虑 $y=y(x)$ 或 $x=x(y)$ 为光滑的微分同胚的情况.这样,这些坐标可以说是光滑的局部坐标.

定理中只泛泛地说 $f$ 是光滑映射,其实是指,如果 $f\in C^k$,$k\geqslant1$,或 $C^\infty$,则 $f$ 作为微分同胚也具有相同的光滑性,即逆映射也是 $C^k$,$k\geqslant1$ 或 $C^\infty$ 的.

定理的证明是利用著名的压缩映像原理.因为0在 $\mathbf{R}^n$ 中之任一邻域中必包含一个球 $B_r(0)$ ($B_r(P)$表示 $P$ 为心,$r$ 为半径的开球体,$S_r(P)$则表示这个球的球面).以下我们恒设所说的邻域是一个球体,球体是一个凸集,即对球内任意两点 $x_1,x_2$,连接它们的线段 $\{tx_1+(1-t)x_2,0\leqslant t\leqslant1\}$ 也全在球内.这就是说,球不但对于球心是星形区域,而且对其中任意点都是星形区域.因此任一可微函数都变成 $t$ 的函数,而可以应用中值定理.

由假设 $(\mathrm{d}f)(0)$ 是非退化的.于是在一定坐标系下它是一个非奇异矩阵 $A$,即映射 $f$ 的雅可比矩阵.现用 $A^{-1}$ 左乘 $f$,而 $A^{-1}f$ 在 $x=0$ 的微分为 $I$.求解 $f(x)=y$ 与求解 $(A^{-1}\circ f)(x)=A^{-1}y$ 是等价的,而 $\mathrm{d}(A^{-1}f)(0)=I$,因此不失一般性,以下我们都设映射 $f$ 适合条件 $(\mathrm{d}f)(0)=I$.

我们下面不考虑映射 $f$,而考虑映射

$$g:\Omega\subset\mathbf{R}^n\to\mathbf{R}^n,x\mapsto f(x)-x\text{(即 }g(x)=f(x)-x\text{)}. \tag{14}$$

它仍然适合 $g(0)=0$,而且

$$(\mathrm{d}g)(0)=(\mathrm{d}f)(0)-I=0. \tag{15}$$

因此，当 $x\in\bar{B}_\rho(0)$ 且 $\rho$ 充分小时，必有 $\|(\mathrm{d}g)(x)\|<\eta$，$\eta$ 是任意小的正数.

现在我们用迭代法去解方程

$$x=y-g(x). \tag{16}$$

但是这个方程的解就是映射 $y-g(x)$ 的不动点：$y-g(\cdot):x\mapsto x$. 因为这个 $x$ 必适合

$$x=y-g(x)=y+x-f(x) \text{或} f(x)=y.$$

求解(16)可以使用古典的逐次逼近法. 即对 0 附近的 $y$，任取 $x_0\in B_\rho(0)$ 作为零次近似. 以 $x_0$ 代入(16)右方，并记所得为 $x_1$：

$$x_1=y-g(x_0)=y+x_0-f(x_0),$$

然后逐次地做下去，即令

$$x_{l+1}=y-g(x_l),\cdots \tag{17}$$

这样得到一个序列 $\{x_l\}$. 我们希望 $\{x_l\}$ 有极限 $\bar{x}$，而且 $\bar{x}$ 即是(16)的不动点.

这样做要解决两个问题：一是要保证每次得到的 $x_l$ 都在 $g(x)$ 之定义域（亦即 $f$ 之定义域）中. 例如适合 $x_l\in B_\rho(0)$. 然后再证 $\{x_l\}$ 收敛. 为了做到这两点，首先是取 $y\in B_{\frac{\rho}{2}}(0)$，$\rho$ 待定，然后证明映射 $g$ 在 $B_\rho(0)$ 中适合

$$\|g(x^*)-g(x^{**})\|\leqslant\frac{1}{2}\|x^*-x^{**}\|. \tag{18}$$

这个式子的证明见后.

于是对于固定的 $y$. 如果已作出了 $x_l$，则对于 $x_{l+1}=y-g(x_l)$，因为 $g(0)=0$，故有

$$\begin{aligned}\|x_{l+1}\|&\leqslant\|y\|+\|g(x_l)-g(0)\|<\frac{\rho}{2}+\frac{1}{2}\|x_l-0\|\\&\leqslant\frac{\rho}{2}+\frac{\rho}{2}=\rho.\end{aligned}$$

即是说，只要 $\|x_l\|<\rho$，必有 $\|x_{l+1}\|<\rho$. 但 $x_0$ 是取在 $B_\rho(0)$ 中的，所以一切 $x_l$ 都在此球中，从而也在 $f(x)$ 与 $g(x)$ 之定义域中.

其次由(18)式

$$\|x_{l+1}-x_l\|=\|g(x_l)-g(x_{l-1})\|\leqslant\frac{1}{2}\|x_l-x_{l-1}\|.$$

即是说每经一次映射 $g$，$x_{l+1}$ 与 $x_l$ 之距离比 $x_l$ 与 $x_{l-1}$ 之距离至少减半——所谓压缩映像即指此. 而反复应用它，有

$$\|x_{l+k}-x_l\|\leqslant\left(\frac{1}{2}\right)^l\|x_k-x_0\|.$$

所以 $\{x_l\}$ 是 $\mathbf{R}^n$ 中的柯西序列，而必有极限 $\bar{x}\in\bar{B}_\rho(0)$. 在(17)中求极限，即得

$$\bar{x}=y-g(\bar{x})=y+\bar{x}-f(\bar{x}),$$

亦即

$$y=f(\bar{x}).$$

从而 $\bar{x}$ 是方程 $y=f(x)$ 对固定于 $\overline{B}_\rho(0)$ 中的 $y$ 的解.

再证这个解是唯一的. 设有 $\bar{\bar{x}}$ 也是 $y=f(x)$ 之解,则 $f(\bar{\bar{x}})=f(\bar{x})$,于是 $g(\bar{\bar{x}})-g(\bar{x})=\bar{\bar{x}}-\bar{x}$. 但由此有

$$\| \bar{\bar{x}}-\bar{x} \| = \| g(\bar{\bar{x}})-g(\bar{x}) \| \leqslant \frac{1}{2} \| \bar{\bar{x}}-\bar{x} \| .$$

所以 $\bar{\bar{x}}=\bar{x}$.

余下的只是证明(18)式了. 在 $B_\rho(0)$ 中作直线段 $\{tx^*+(1-t)x^{**}, 0\leqslant t\leqslant 1\}$ 连接 $x^*$ 与 $x^{**}$ 两点. 在这个直线段上 $g$ 成为 $t$ 的光滑函数,从而

$$\begin{aligned} g(x^*)-g(x^{**}) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} g(tx^*+(1-t)x^{**})\mathrm{d}t \\ &= \int_0^1 (x^*-x^{**})(\mathrm{d}g)(tx^*+(1-t)x^{**})\mathrm{d}t . \end{aligned}$$

由于在 $\bar{B}_\rho(0)$ 中 $\| \mathrm{d}g \| < \eta$, $\eta$ 是任意小数,故当 $\rho$ 充分小时

$$\| g(x^*)-g(x^{**}) \| \leqslant \frac{1}{2} \| x^*-x^{**} \| .$$

这就是(18)式.

现在证明 $f^{-1}$ 也是可微的. 事实上,由 $f$ 为连续可微,可设 $x=x_1+h$ 应有

$$f(x)=f(x_1)+(\mathrm{d}f)(x_1)h+o(h). \tag{19}$$

这里 $(\mathrm{d}f)(x_1)$ 当 $x_1$ 在 0 附近时,由假设是非退化的. 用 $A(x_1)$ 表示 $(\mathrm{d}f)(x_1)$,则 $A^{-1}(x_1)$ 是存在的. 用 $A^{-1}(x_1)$ 作用于上式两侧有

$$A^{-1}(x_1)[f(x)-f(x_1)]=x-x_1+A^{-1}(x_1)o(x-x_1).$$

若记 $y=f(x)$, $y_1=f(x_1)$,则 $x_1=f^{-1}(y_1)$,(19)式可以写为

$$f^{-1}(y)=f^{-1}(y_1)+A^{-1}(x_1)(y-y_1)-A^{-1}(x_1)o(x-x_1). \tag{20}$$

不过由前面的假设(即已用 $A^{-1}(0)$ 遍"乘"方程 $y=f(x)$ 双方),可以设 $A(0)=I$. 所以当 $x, x_1\in B_\rho(0)$ 从而 $y, y_1\in B_{\frac{\rho}{2}}(0)$ ($\rho$ 已设为充分小),有两个正常数 $\lambda$ 与 $\Lambda$ 存在使

$$0<\lambda\leqslant\frac{\| x-x_1 \|}{\| y-y_1 \|}\leqslant\Lambda .$$

这样知道 $o(x-x_1)$ 即 $o(y-y_1)$. 代入(20),注意到 $\| A^{-1}(x_1) \| \leqslant M$ ($M$ 是一个适当正数),我们有(记 $y-y_1=k$)

$$f^{-1}(y)=f^{-1}(y_1)+A^{-1}(x_1)k+o(k). \tag{21}$$

这就是说 $f^{-1}(y)$ 在 $y_1$ 处连续可微. 但是 $y_1=f(x_1)$ 可以选为 $B_{\frac{\rho}{2}}(0)$ 中任一点,所以 $f^{-1}$ 在 $y=0$ 附近连续可微.

定理中是说逆映射 $f^{-1}$ 与 $f$ 有相同的光滑程度. 但这只要利用泰勒公式即可得证. 证明的大意如下. 如果 $f(x)$ 在 $x=0$ 附近属于 $C^l$,则在 $B_\rho(0)$ ($\rho$ 充分小)中任取一点 $x_1$,并记 $x=x_1+h$,这里 $h=(h_1,\cdots,h_n)$ 是一个 $n$ 维向量,我们应有

$$f(x)=f(x_1)+A(x_1)(x-x_1)+\sum_{|\alpha|=2}^{l} A_\alpha(x_1)h^\alpha+o(h^{l+1}).$$

这里 $h=x-x_1=f^{-1}(y)-f^{-1}(y_1)$. 令

$$x-x_1=B(y_1)(y-y_1)+\sum_{|\beta|=2}^{l}B_\beta(y_1)(y-y_1)^\beta+O(\|y-y_1\|^{l+1}),\tag{22}$$

以它代入上式. 注意到我们已经得出 $B(y_1)=A^{-1}(x_1)$. 现用待定系数法计算 $B_\beta(y_1)$，注意到为此只需利用 $A^{-1}(x_1)$即可，而不必考虑其余的 $A_\alpha(x_1)$是否为退化. 于是即可得到(22)式而知逆映射 $x=f^{-1}(y)$在 $y=0$ 附近，至少在 $B_{\frac{\varrho}{2}}(0)$中是 $C^l$ 可微的.

以上我们全是对实变量 $x,y$ 来证明反函数定理的. 但是这个证明过程对复的自变量仍然适用. 这一点读者容易看到. 所以若定理中的 $f$ 是 $n$ 个复变量 $z=(z_1,\cdots,z_n)$到其他 $n$ 个复变量 $w=(w_1,\cdots,w_n)$的映射，$f(0)=0$，而且 $f(z)$在 $z=0$ 的邻域中对 $z$ 是解析的(上节中已说了，对复变量的连续可微性就是解析性，这一点对多个复变量也成立，而且所谓对 $z=(z_1,\cdots,z_n)$为解析就是对每一个 $z_i$ 分别为解析)，则只要$(\mathrm{d}f)(0)$是非退化的，$w=f(z)$一定有唯一解析的逆映射 $z=f^{-1}(w)$存在. 这里一切都是在 $z=0$ 与 $w=0$ 的邻域中来讲的.

**3. 单射定理与投影定理**　下面我们讨论 $m$ 与 $n$ 不相等的情况. 我们将得出这样的结论：前面关于线性映射的结论，当 rank $A$ 取最大可能的值时，局部地都适用于一般的映射 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$，这里 $f$ 是适合 $f(0)=0$ 的光滑映射，而 $\Omega$ 是 0 的充分小邻域. 上面关于映射光滑性的说明完全适用于此，而且 $\Omega$ 是充分小邻域这句话我们也不再重复了.

首先看 $m<n$ 的情况. 直观地看，$f$ 是把 $\mathbf{R}^m$ 中的 $\Omega$ 映为 $\mathbf{R}^n$ 中过 0 的"曲面"$f(\Omega)$. 如果$(\mathrm{d}f)(0)$是非退化的，则其秩是 $m$，"曲面"$f(\Omega)$在 $f(0)=0$ 附近可否看成一个"平面"，即可否用它在$x=0$处的切平面代替它？到现在为止我们的经验告诉我们应该是可以的. 因为$(\mathrm{d}f)(0)$具有最大可能的秩，在 $m=1$ 时情况(这时 $f(\Omega)$是曲线)$x(t)=(x_1,\cdots,x_n)$，最大可能秩为 $m=1$ 就相当于 $\dot{x}(t)\neq 0$，亦即曲线 $x=x(t)$上没有奇点. 因此现在我们可以设想，$(\mathrm{d}f)(0)$具有最大可能秩 $m$，也就意味着"曲面"$f(\Omega)$上没有奇点(没有临界点). 但是我们还要再提醒一下，这里奇点并不是某一坐标或其某阶导数变成无穷之类，而这也是不会有的，因为我们已经假设了 $f$ 是光滑的. 读者可以回想一下莫尔斯引理，那里是讲的一个函数(但是它是 $m>1$ 而 $n=1$ 的情况，与此不同)，化成 $f=\sum_{i=1}^{\lambda}x_i^2-\sum_{j=1}^{m-\lambda}x_{\lambda+j}^2$ 后，说 $x=0$ 是一个奇点——临界点，即指该函数的所有的一阶偏导数在 $x=0$ 时均为 0，因此有许多事就不能用 $f$ 来作. 例如不可以作一个坐标变换，使 $f=0$ 变成"坐标平面"，比如 $y_1=0$. 这是因为上面我们已经说了，所谓坐标变换必须是至少局部地是微分同胚，因此其雅可比行列式不能为 0. 但若以 $y_1=f$，则雅可比行列式有一行在 $x=0$ 处全为 0：

$$\left(\frac{\partial y_1}{\partial x_1},\cdots,\frac{\partial y_1}{\partial x_m}\right)_{x=0}=\left(\frac{\partial f_1}{\partial x_1},\cdots,\frac{\partial f_1}{\partial x_m}\right)_{x=0}=0.$$

而我们现在正是想作一个坐标变换 $g$，把曲面 $f(\Omega)$在原点附近"拉平"成为"坐标平面"(见图 3-5-3右列). 如果做到了这一步，从几何上看，也就与图 3-5-1 完全一样了. 可见现在的坐标变换 $g$ 就起了一个"拉平"、"拉直"的作用. 那么这个 $g$ 与原来的$f$ 究竟是何关系呢？为了弄清这一点，我们把源空间"添上"一个 $\mathbf{R}^{n-m}$维子空间(用线性代数语言来说就是求"直和")成为一个扩大的源空间 $\mathbf{R}^n$ 而与靶空间同维数. 图 3-5-3 左用虚线描述了这个过程. 这时，原来的映射也"扩大"成了 $F$. 于是整个过程就是：通过 $F$ 把$\Omega$"变曲"为靶空间中的 $f(\Omega)$. 再在靶空间中作

坐标变换 $g$ 把 $f(\Omega)$ 拉平. 总之,顺着图 3-5-3 的斜线看,我们有一个由扩大的源空间 $\mathbf{R}^n$ 到作为靶空间的 $\mathbf{R}^n$ 的映射,它是由 $g\circ F$ 实现的. 如果 $g\circ F = id$,则 $g$ 必适合所求. 所以 $g$ 是 $F$ 的逆映射. 这样一来,我们会得到一个什么样的定理,又如何去证明它也就十分清楚了. 定理 3 的记号与前面讲的一样.

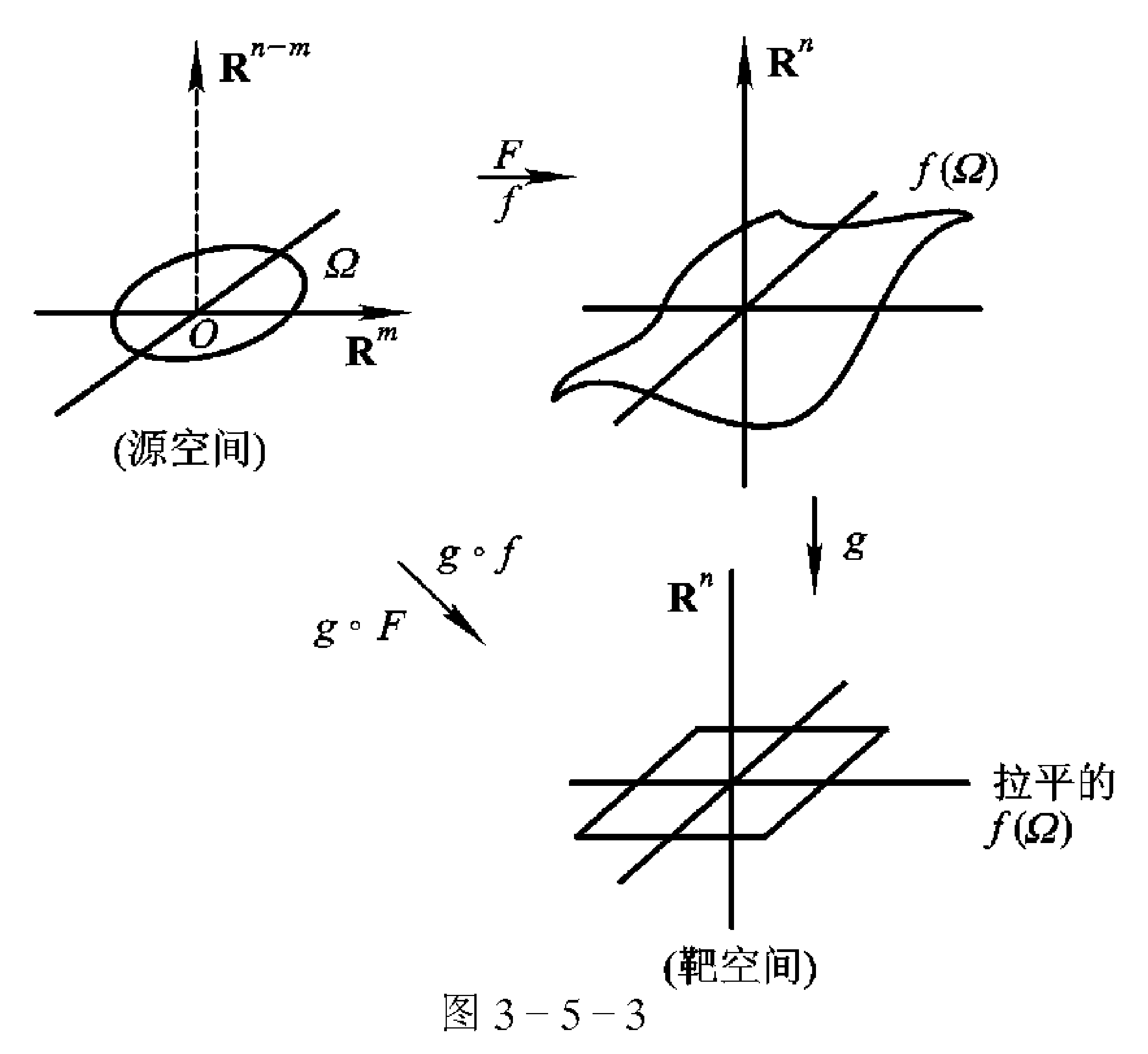



图 3-5-3

**定理 3(单射定理)** 设 $m\leqslant n$. $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$ 是一个光滑映射,且 $f(0)=0$. 若 $(\mathrm{d}f)(x)$ 在 $x=0$ 附近恒取最大可能秩,则必有由 $0\in\mathbf{R}^n$ 的某个充分小邻域(即设为 $\Omega$)到 0 的另一个邻域的微分同胚 $g$,使得 $g(0)=0$,而且

$$(g\circ f)(x_1,\cdots,x_m)=\begin{pmatrix}x_1\\ \\ x_m\\ 0\\ \\ 0\end{pmatrix}\qquad(\text{后 } n-m \text{ 个元为 } 0). \tag{23}$$

**证** 不妨设 $f$ 的雅可比矩阵的左上角子矩阵

$$J_1=\left(\frac{\partial f_i}{\partial x_j}\right)_{1\leqslant i,j\leqslant m}$$

在 $x=0$ 附近是非退化的. "扩大"源空间,即引入新的 $\mathbf{R}^{n-m}$,令其中之点的坐标是 $(x_{m+1},\cdots,x_n)$,并定义新的映射:

$$F(x_1,\cdots,x_n)={}^t(f_1(x_1,\cdots,x_m),\cdots,f_n(x_1,\cdots,x_m))+{}^t(0,\cdots,0;x_{m+1},\cdots,x_n):\ \mathbf{R}^n\to\mathbf{R}^n. \tag{24}$$

显然 $F(0)=0$,而且它在 $x=0$ 处的雅可比矩阵是

$$J(0)=\begin{pmatrix}\dfrac{\partial f_i}{\partial x_j}(0) & 0\\ * & I\end{pmatrix}=\begin{pmatrix}J_1(0) & 0\\ * & I\end{pmatrix}.$$

由于已设 $\det J_1(0)\neq 0$，所以 $\det J(0)\neq 0$，而由反函数定理. $F$ 必有局部逆 $F^{-1}=g, g(0)=0$，而 $g\circ F=id$: $\mathbf{R}^n\to\mathbf{R}^n$. $\mathbf{R}^n$ 既是扩大的源空间，又是靶空间. 把 $g$ 用局部坐标写成 $g_i=g_i(y_1,\cdots,y_n)$，$1\leqslant i\leqslant n$，则 $g\circ F=id$. 即

$$g_1(F(x_1,\cdots,x_n))=x_1,\tag{25}$$
$$\cdots\cdots\cdots\cdots$$
$$g_n(F(x_1,\cdots,x_n))=x_n.$$

再由扩大的源空间回到原来的源空间，即令 $x_{m+1}=\cdots=x_n=0$. 由 (24) 知这时 $F_i$ 变成 $f_i$，而 (25) 即成为 (23). 定理证毕.

**注**　我们不说 $(\mathrm{d}f)(x)$ 在 $x=0$ 处之秩为 $m$，却说 $(\mathrm{d}f)(x)$ 在 $x=0$ 附近的秩均为 $m$，这是因为秩是一个可以突然变化的量. 例如二阶矩阵 $\begin{pmatrix}x & 0\\ 0 & 1\end{pmatrix}$ 在 $x\neq 0$ 处的秩是 2，而在 $x=0$ 时却突然变为 1. 这个情况正是我们想要避免的. 本来，假设了 $f$ 光滑，至少 $f\in C^1$，则 $\operatorname{rank} A=\operatorname{rank}(\mathrm{d}f)(x)$ 是 $x$ 的连续函数. 若 $\operatorname{rank}(\mathrm{d}f)(0)=m$，则在 $x=0$ 附近均有 $\operatorname{rank}(\mathrm{d}f)(x)=m$，而上面的情况不会发生. (但若 $\operatorname{rank} A$ 在某点小于最大可能值 $m$，在其附近则不一定不发生突变了. 这种情况我们的说法是：非退化性是稳定的，而退化性则是不稳定的). 我们对定理的陈述在“$x=0$”后面画蛇添足地加了“附近”二字，无非是提醒读者注意这件事.

下面我们再来看 $m>n$ 的情况. 我们仍从几何说明开始，现在情况就简单多了. 在映射 $f$ 下，靶空间中的原点 0 之原像 $f^{-1}(0)$ 是源空间中过 $x=0$ 的一个“曲面”. 其维数是 $m-n$. 图 3-5-4中画的是一条曲线，这是因为更高维数的几何图形画不出来. 问题在于它会不会有奇点? 在讲单射定理时，我们以 $m=1$ 的特例 (这时映射 $f$ 成了一条曲线)，比照着曲线上有没有临界点的情况，说 $f(\Omega)$ 是 $\mathbf{R}^n$ 中的曲面. 现在不妨也看一个特例，设 $n=2$，于是映射 $f$ 成为

$$y_1=f_1(x_1,x_2,\cdots,x_m),$$
$$y_2=f_2(x_1,x_2,\cdots,x_m).\tag{26}$$

而 $f^{-1}(0)$ 就是将上式左方的 $y_1,y_2$ 换成 0，并求其全部解而得. 例如设可从 $f_1=0, f_2=0$ 中解出 $x_1,x_2$:

$$x_1=\varphi_1(x_3,x_4,\cdots x_m),$$
$$x_2=\varphi_2(x_3,x_4,\cdots,x_m)\tag{27}$$

(下面我们会看到，当“最大秩”假设 $\operatorname{rank}\mathrm{d}f(0)=n$ 成立时这总是可以作到的)，由此可见 $f^{-1}(0)$ 是一个 $m-n$ 维“曲面”，它是没有“奇点”的，而 (27) 是它的参数表示. 到了这一步，下面就应考虑如何把 $f^{-1}$“拉平”. 办法仍是作坐标变换. 不过现在要在源空间中作变换了，这就是图3-5-4 左侧，由下到上是 $h$，$h$ 当然不会是线性的，否则它不会把弯曲的 $f^{-1}(0)$ 拉直. 但是再加上一个向右的非线性的 $f$，却成为由直的源空间 $\mathbf{R}^n$ 到直的靶空间 $\mathbf{R}^n$ 的 $f\circ h$. 我们想到是否可以通过“扩大”靶空间，把两个非线性映射复合成一个线性的投影映射 $\pi:\mathbf{R}^m\to\mathbf{R}^n$? 办法和上面是一样的：“扩大”靶空间为 $\mathbf{R}^m$，同时“扩大”映射为 $F$，使 $F$ 成为一个微分同胚，再作其逆 $h$，然后再限制 $F$ 为 $f$ 即可.

**定理 4 (投影定理)**　设 $f:\Omega\subset\mathbf{R}^m\to\mathbf{R}^n$，是一光滑映射，$\Omega$ 是 0 的一个充分小邻域，$f(0)=0$. $m\geqslant n$，若 $(\mathrm{d}f)(x)$ 在 $x$ 附近恒取最大可能秩 $n$，则必有由 $\Omega$ 到 0 在 $\mathbf{R}^m$ 中的另一邻域的微分

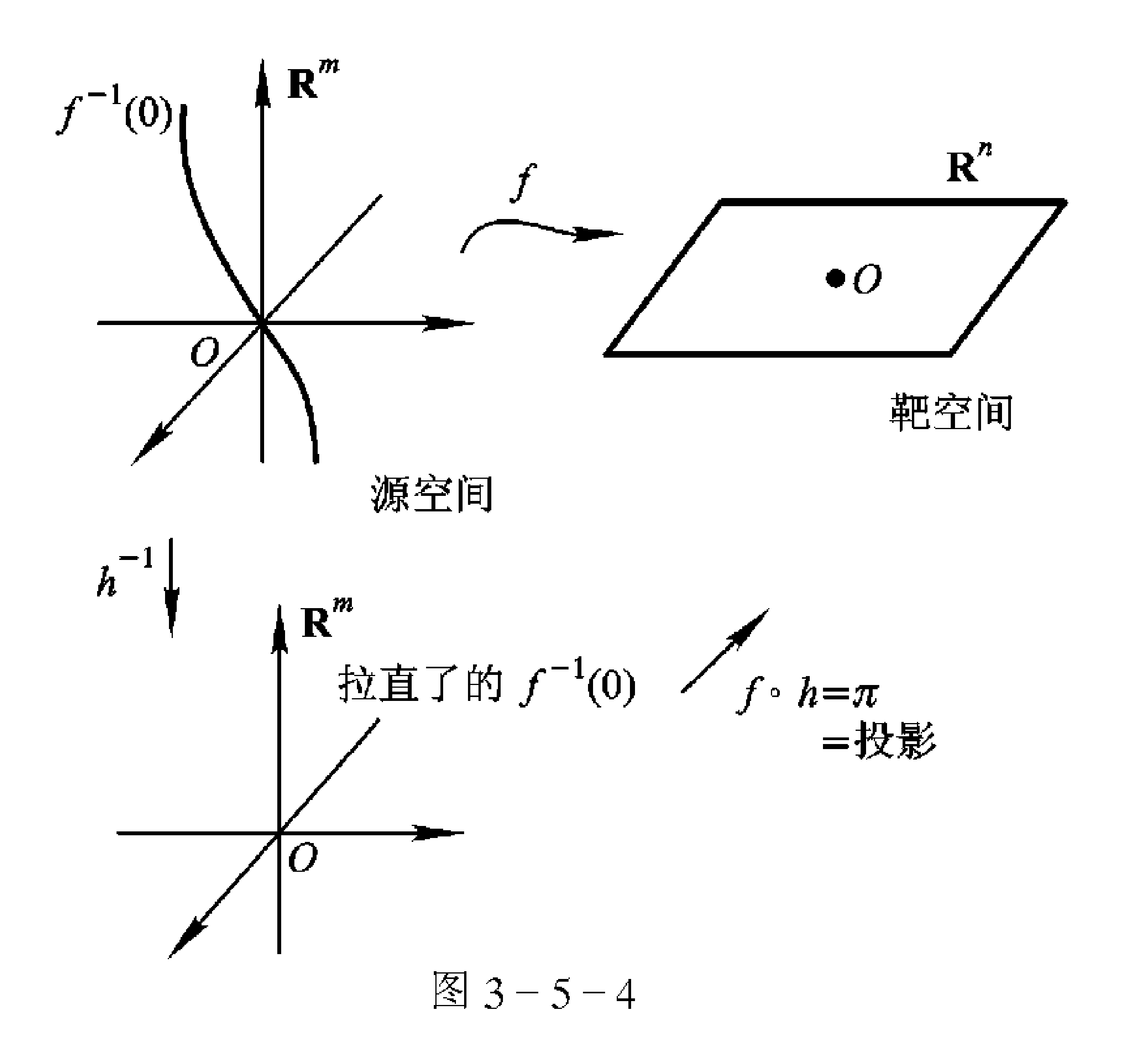



图 3-5-4

同胚 $h$, $h(0)=0$,使得

$$(f\circ h)(x_1,\cdots,x_m)=\pi(x_1,\cdots,x_{m-n};x_{m-n+1},\cdots,x_m)=(x_{m-n+1},\cdots,x_m). \tag{28}$$

它将 $f^{-1}(0)=\{(x_1,\cdots,x_{m-n};0,\cdots,0)\}$零化.

**证**　$(\mathrm{d}f)(x)$是一个 $n\times m$ 矩阵,

$$J=\begin{pmatrix}\dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_m}\\ \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \dfrac{\partial f_n}{\partial x_m}\end{pmatrix},$$

设它的后 $m-n$ 列所成的子矩阵

$$J_1=\begin{pmatrix}\dfrac{\partial f_1}{\partial x_{m-n+1}} & \cdots & \dfrac{\partial f_1}{\partial x_m}\\ \\ \dfrac{\partial f_n}{\partial x_{m-n+1}} & \cdots & \dfrac{\partial f_n}{\partial x_m}\end{pmatrix}$$

非退化.现在把靶空间扩大为 $m$ 维空间(在每个向量前面加 $m-n$ 个坐标$(y_1,\cdots,y_{m-n})$).同时把 $f$ 扩大为

$$F(x_1,\cdots,x_n)=\begin{pmatrix}x_1\\ \\ x_{m-n}\\ f_1(x)\\ \\ f_n(x)\end{pmatrix},$$

即成为

$$y_i = F_i(x) = x_i, 1 \leqslant i \leqslant m-n,$$
$$y_j = F_j(x) = f_j(x), m-n+1 \leqslant j \leqslant m.$$

容易看到 $F$ 对 $x$ 之雅可比矩阵是

$$\begin{pmatrix} I & 0 \\ * & J_1 \end{pmatrix},$$

它在 $x=0$ 附近是非退化的，而由反函数定理知 $F$ 在 0 附近有局部逆 $h$，令 $\pi$ 是由 $\mathbf{R}^m$ 到 $\mathbf{R}^n$ 的投影线性映射，$\pi: \mathbf{R}^m \to \mathbf{R}^n$，$(x_1, \cdots, x_{m-n}, x_{m-n+1}, \cdots, x_n) \mapsto (x_{m-n+1}, \cdots, x_m)$，则 $f = \pi \circ F$. 从而

$$(f \circ h)(x_1, \cdots, x_m) = (\pi \circ F \circ h)(x_1, \cdots, x_m) = \pi(x_1, \cdots, x_m) = (x_{m-n+1}, \cdots, x_m).$$

定理证毕.

由于定理 2,3,4 极为重要，我们要多花一些功夫加以解释. 定理 3,4 都涉及"曲面"，但是什么是曲面我们除了在古典的微积分教本中有不甚明确的描述以外，还没有认真讨论过. 这要等到第七章才知道，它们都是某个欧氏空间中的子流形. 但现在我们就已看到，确定一个子流形有两个方法. 其一是 $\mathbf{R}^n$ 中的 $m$ 维子流形 $(m<n)$，局部也都是 $\mathbf{R}^m$ 中某点的邻域 $\Omega$ 的像，所以可以局部地表示为

$$y_i = f_i(x_1, \cdots, x_m), i = 1, 2, \cdots, n, m < n, \tag{29}$$

这里 $f_i$ 是光滑函数. 而对于 $f_i$ 所构成的映射 $f: \Omega \subset \mathbf{R}^m \to \mathbf{R}^n$，要求局部地适合最大秩条件，即 rank $(\mathrm{d}f)(x) = m$. 但是 $(\mathrm{d}f)(x)$ 即上述映射的雅可比矩阵，所以，最大秩条件现在是：存在某一个不为 0 的子行列式.

$$\det \begin{pmatrix} \dfrac{\partial f_{i_1}}{\partial x_1} & \cdots & \dfrac{\partial f_{i_1}}{\partial x_m} \\ & \cdots & \\ \dfrac{\partial f_{i_m}}{\partial x_1} & \cdots & \dfrac{\partial f_{i_m}}{\partial x_m} \end{pmatrix} \neq 0. \tag{30}$$

(30)表示(29)中有 $m$ 个函数 $\{f_i\}$ 是互相独立的，即不会有例如

$$f_{i_1} = \Phi(f_{i_2}, \cdots, f_{i_m})$$

的情况，否则，(29)式形式上是 $n$ 个函数实际上独立的会少于 $n$ 个. (29)式就是通常微积分教本中讲的曲面的参数表示. 源空间其实就是参数 $(x_1, \cdots, x_m)$ 所成的空间.

表示曲面的另一个方法是给出定义它的方程

$$f_i(x_1, \cdots, x_m) = 0, \quad i = 1, 2, \cdots, n. \tag{31}$$

但是这一组方程的左方函数又定义了一个映射 $f = (f_1, \cdots, f_n): \Omega \subset \mathbf{R}^m \to \mathbf{R}^n, m > n$,

$$y_i = f_i(x_1, \cdots, x_m).$$

所以这个曲面就是上述映射下 $0 \in \mathbf{R}^n$ 的原像 $f^{-1}(0)$. 这时最大秩条件 rank $(\mathrm{d}f)(x) = n$ 又是不可少的. 否则(31)中实际上独立的个数少于 $n$. 当(31)中独立方程个数为 $n$ 时，它们局部地定义一个 $m-n$ 维曲面. 若独立方程个数小于 $n$，则它定义一个更高维的对象.

作为曲面，不论用什么方法去确定它都不应该影响它的性质. 那么曲面的性质是什么呢？这

两个定理告诉我们：就是局部地可以“拉平”，也就是微分同胚于一个线性空间的开集. 前一个情况下是同胚于源空间 $\mathbf{R}^m$ 中 0 的邻域 $\Omega$. 后一个情况我们没有说. 实际上是同胚于 $\mathbf{R}^{m-n}$ 中 0 的某个邻域 $\Omega$. 我们前面多次说这个曲面是无奇点的. 那么，什么是无奇点呢？现在可以明确了，无奇点就是微分同胚于某个 $\mathbf{R}^m$ 或 $\mathbf{R}^{m-n}$ 中的开集，最大秩条件才能保证无奇点.

定理 2，3，4 都把几何问题化成了代数问题. 试以定理 3 为例，那里的证明是通过构造 $f$（其实是 $F$）之左逆 $g$ 来实现的. 不过要注意，这里不能写为 $f\circ g$，因为 $g:\mathbf{R}^n\to\mathbf{R}^n$ 是扩大了的源空间中的微分同胚. $g$ 的像在 $\mathbf{R}^n$ 中，而 $f$ 只定义在 $\mathbf{R}^m$ 中，所以 $f\circ g$ 是没有意义的. 因此，定理 3 是通过作 $f$ 的左逆来证明的. 同样，定理 4 是用的 $f$ 的右逆 $h$. 因此定理 3 是说 $f$ 有左逆 $g$：$g\circ f=id$.（不过这个记号不准确. 因为 $g\circ f:\mathbf{R}^m\to\mathbf{R}^n$，而恒等算子 $id$ 不会作用于不同空间中.）用链式法则有

$$\mathrm{d}g\circ\mathrm{d}f=\begin{pmatrix}I\\0\end{pmatrix}.$$

这里 $\mathrm{d}f$ 是 $n\times m$ 矩阵. 其左逆 $\mathrm{d}g$ 是 $n\times n$ 矩阵，这样上式才有意义. 类似地，定理 4 是说 $f$ 有右逆 $h$：$f\circ h=id$，故

$$\mathrm{d}f\circ\mathrm{d}h=(0\quad I).$$

$\mathrm{d}h$ 是 $m\times m$ 矩阵. 现在我们再回到线性代数的问题，设有 $n\times m$ 矩阵 $A$，问

$$Ax=b\qquad b\in\mathbf{R}^n,x\in\mathbf{R}^m\tag{32}$$

何时有解？何时解为唯一的？现在很清楚，若 $A$ 有左逆 $A_l^{-1}$，则 (32) 若有解必是唯一的：

$$x=A_l^{-1}b.$$

这时 $A$ 必为单射. 类似地，若 $A$ 有右逆 $A_r^{-1}$，则 (32) 必可解，因为 $x=A_r^{-1}b$ 就是 (32) 的解. 这时 $A$ 必为满射，也就是投影线性变换. 总之我们可以列表如下：

| | $m<n$ | $m=n$ | $m>n$ |
|---|---|---|---|
| 最大秩条件 | $\mathrm{rank}(\mathrm{d}f)(x)=m$ | $\mathrm{rank}\ (\mathrm{d}f)(x)=m=n$ | $\mathrm{rank}\ (\mathrm{d}f)=n$ |
| 逆算子情况 | $(\mathrm{d}f)(x)$有左逆 | $(\mathrm{d}f)(x)$可逆 | $(\mathrm{d}f)(x)$有右逆 |
| 映射的特点 | $(\mathrm{d}f)(x)$为单射 | $(\mathrm{d}f)(x)$为一一映射 | $(\mathrm{d}f)(x)$为满射 |
| 几何特性 | $f$ 为“浸入” | $f$ 为微分同胚 | $f$ 为“外罩” |
| 子流形 | $f(\Omega)$为 $m$ 维子流形 | $f(\Omega)$为 $m=n$ 维子流形，即 0 之另一个邻域 | $f^{-1}(0)$为 $m-n$ 维子流形 |
| 结论 | 定理 3 | 定理 2 | 定理 4 |

我们前面已说过 $\mathrm{d}f$ 即相应于 $f$ 的切映射. 它是线性空间之间的映射，而这些线性空间是相应流形（$m<n$ 时的 $f(\Omega)$：$m=n$ 时的 $\Omega$；$m>n$ 时的 $f^{-1}(0)$）的切空间. 所以，如果用一句话来概括以上所说，那就是：在最大秩条件下，切空间之间的切映射，很好地局部表现了原空间之间的映射的几何特性.

我们不妨再从一本名著 L. Hörmander. The Analysis of Linear Partial Differential Operators, $I$ 中抄录一段（文字稍有修改）看一下该书中这套定理总结成什么：

**定理** 设 $\Omega\subset\mathbf{R}^m$，$f:\Omega\to\mathbf{R}^n$ 为 $C^1$ 映射，$0\in\Omega$，且 $f(0)=0$. 存在由 $0=f(0)$ 在 $\mathbf{R}^n$ 中的一个

邻域 $\omega$ 到 $\mathbf{R}^m$ 的 $C^1$ 映射 $g:\omega\to\mathbf{R}^m$ 使得 a) $f\circ g=id$ 于 $f(0)$ 附近，即 $g$ 为 $f$ 之局部右逆，b) $g\circ f=id$ 于 0 附近，即 $g$ 为 $f$ 之局部左逆；c) $g\circ f=id$，于 0 附近，且 $f\circ g=id$ 于 $f(0)$ 附近，即 $g$ 为 $f$ 之局部逆．其充要条件是存在一个线性映射 $A:\mathbf{R}^n\to\mathbf{R}^m$ 使得 a′) $(\mathrm{d}f)(0)A=id_{\mathbf{R}^n}$；b′) $A(\mathrm{d}f)(0)=(id_{\mathbf{R}^m}$；c′) $(\mathrm{d}f)(0)A=id_{\mathbf{R}^n}$ 同时 $A(\mathrm{d}f)(0)=id_{\mathbf{R}^m}$．a′) 等价于 $(\mathrm{d}f)(0)$ 为满射；b′) 等价于 $(\mathrm{d}f)(0)$ 为单射；c′) 等价于 $(\mathrm{d}f)(0)$ 为一一映射，且 $g$ 在 $f(0)=0$ 附近是唯一的．

证．a′)，b′) 与 c′) 之必要性均可由链式法则直接得出．所以下面只证明它们的充分性． c) 中 $g$ 的唯一性可以从以下事实看到：若 $g_1$ 与 $g_2$ 是存在的．$f\circ g_1=id$，$g_2\circ f=id$ 已经得证．则 $g_1=g_2\circ f\circ g_1=g_2\circ(f\circ g_1)=g_2\circ id=g_2$．故所有左逆 $g_2$ 均等于某一指定的右逆 $g_1$，从而左逆唯一．同理右逆也唯一．因此，定理归结为由 a′)、b′) 证明 $g_1$、$g_2$ 存在．这可由逐步逼近法完成．今设 a′) 或 b′) 成立，若必要时用 $(\mathrm{d}f)(0)$ 去左乘或右乘 $A$，可以假设 $(\mathrm{d}f)(0)=I$．取 $\delta>0$ 充分小可以设当 $\|x\|<\delta$ 时 $\|\mathrm{d}f(x)-I\|<\frac{1}{2}$．因此，若 $x_1,x_2$ 均在 $B_{\frac{\delta}{2}}(0)$ 内，由 $f$ 之可微性，令 $h=x_1-x_2$，有

$$\|f(x_1)-f(x_2)-I(x_1-x_2)\|\leqslant\frac{1}{2}\|x_1-x_2\|. \tag{33}$$

故若 $f(x_1)=f(x_2)$，则上式成为 $\|x_1-x_2\|\leqslant\frac{1}{2}\|x_1-x_2\|$，而当 $x_1\neq x_2$ 时这就是矛盾．因此，若 a′)，b′) 成立，则 a)，b) 中的 $g$ 若存在必为唯一的．现在只需证明对任一个 $y\in B_{\frac{\delta}{2}}(0)$，一定在 $B_{\frac{\delta}{2}}(0)$ 中能找到一个 $x$ 使 $f(x)=y$．这样一来，$B_{\frac{\delta}{2}}(0)$ 中任一点 $y$ 必对应于 $B_{\frac{\delta}{2}}(0)$ 中唯一一点 $x$ 使 $f(x)=y$．令 $x=g(y)$ 即知局部的左逆或右逆 $g$ 是存在的．从而 a)，b) 中的结论，除 $g\in C^1$ 以外均得证明．所以现证适合 $f(x)=y$ 的 $x$ 在 $B_{\frac{\delta}{2}}(0)$ 中存在．作一串 $\{x_k\}$ 使

$$x_k=x_{k-1}+y-f(x_{k-1}),\ k=1,2,\cdots, \tag{34}$$

其中 $x_0=0$．于是

$$\|x_1-x_0\|=\|x_1\|=\|y\|<\frac{\delta}{2}.$$

如果已证得 $\|x_j\|<\delta$，$j=0,1,\cdots,k-1$，则

$$\begin{aligned}\|x_k-x_{k-1}\|&=\|x_{k-1}-f(x_{k-1})-x_{k-2}+f(x_{k-2})\|\\&\leqslant\frac{1}{2}\|x_{k-1}-x_{k-2}\|\leqslant\cdots\leqslant\frac{1}{2^{k-1}}\|x_1\|\leqslant\frac{\delta}{2^k}.\end{aligned}$$

因此一方面有

$$\|x_k\|\leqslant\|x_k-x_{k-1}\|+\|x_{k-1}-x_{k-2}\|+\cdots+\|x_1-x_0\|<\delta,$$

一方面又有 $\{x_k\}$ 是一个柯西序列，所以有极限 $x\in\overline{B_{\frac{\delta}{2}}}(0)\subset B_\delta(0)$．在 (34) 中求极限即知 $f(x)=y$．

余下只需证明 $g\in C^1$．现在 $g(y)$ 已定义于 $B_{\frac{\delta}{2}}(0)$ 中，令

$$g(y)=x,\ g(y+k)=x+h,$$

则

$$k=f(x+h)-f(x)=(\mathrm{d}f)(x)\cdot h+o(\|h\|).$$

但由(33)，$\|k-h\|<\|h/2\|$. 所以 $\|h\|/2\leqslant\|k\|<2\|h\|$. 而由上式可得，当 $k=0$ 时必有 $h=0$，

$$\|(\mathrm{d}f)(x)h\|=\|k-o(\|h\|)\|>(\frac{1}{2}-\eta)\|h\|,$$

这里 $\eta$ 是任意小正数，因此 $(\mathrm{d}f)(x)$ 可逆. 其逆适合不等式 $\|[(\mathrm{d}f)(x)]^{-1}\|<3$. 所以

$$h=[(\mathrm{d}f)(x)]^{-1}\cdot k+o(k)$$

而知 $g$ 是 $C^1$ 的：

$$(\mathrm{d}g)(y)=[(\mathrm{d}f)(x)]^{-1}, y=f(x).$$

以上的证明录自原书第二版，9 页，定理 1.1.7.

至此我们看到微分学的基本思想——略去高阶无穷小量而只留下 $\Delta f$ 的线性部分 $\mathrm{d}f$，亦即线性化，经过很长时间的发展，吸收了许多新的数学思想和新方法、新概念，例如逐步逼近法、切空间、切映射等许多线性代数的知识凝成几个定理，其内容何等丰富. 然后经大手笔，短短一两页又都写清楚了. 这个总结性的定理我们花了十几页篇幅才讲清楚，原来只是如此. 再仔细一看这个证明没有哪一部分不是我们已经讲过的. 整个数学科学就是这样发展的，我们学好一门数学分支也得这样反复消化. 读者可能会感到难，把多少年的数学发展结晶为这样简洁的证明自非易事. 但是从另一方面看，在"$h$ 是不是无穷小量?"这类云遮雾掩的问题里转，到一两百年才换来今日的满目清明，岂非更难?

**4. 隐函数定理** 隐函数问题在整个数学中具有特殊的重要性. 在非线性分析中它起特别重要的作用. 我们现在要利用反函数定理解决在有限维线性空间中的隐函数问题，它将为更一般的情况提供一个最简单的模型. 这就是设有 $\mathbf{R}^m$ 和 $\mathbf{R}^n$，$x=(x_1,\cdots,x_m)\in\mathbf{R}^m$，$y=(y_1,\cdots,y_n)\in\mathbf{R}^n$. 设在 $U\times V(U\subset\mathbf{R}^m, V\in\mathbf{R}^n)$ 中给出 $n$ 个方程

$$f_i(x_1,\cdots,x_m;\ y_1,\cdots,y_n)=0, i=1,2,\cdots,n \tag{35}$$

而且设 $0\in U$，$0\in V$ 使得 $f_i(0,0)=0$. 问能否从(35)中决定 $y_1\cdots,y_n$ 为 $x$ 的函数：$y_i=y_i(x_1,\cdots,x_m)$，而且适合 $0=y_i(0)$? 如果简单地看，这 $n$ 个 $f_i$ 决定了一个映射

$$f=(f_1,\cdots,f_n)(x,y),$$

它映 $U\times V\to\mathbf{R}^n$. 如果用 $w=(w_1,\cdots w_n)$ 记靶空间 $\mathbf{R}^n$ 中的点. 则求解(35)就变成了研究原像 $f^{-1}(0)$. 但是简单地应用反函数定理是不行的，因为我们要解决的问题是把 $x$ 与 $y$ 分开，它要求在 $f^{-1}(0)$ 上给定了一个 $x$ 恰好有一个 $y=y(x)$，即 $f^{-1}(0)$ 是某个函数的图，而且要求此函数有一定的光滑性. 所以我们要把反函数存在定理的应用稍加修改，而有

**定理 5**(隐函数存在定理) 设 $U\subset\mathbf{R}^m$，$V\subset\mathbf{R}^n$ 是两个开集，其中之点的坐标分别是 $x=(x_1,\cdots,x_m)$，$y=(y_1,\cdots,y_n)$，而且 $0\in U$，$0\in V$. 若 $f: U\times V\subset\mathbf{R}^m\times\mathbf{R}^n\to\mathbf{R}^n$ 是一个光滑映射. $f(0,0)=0$，而且 $(\mathrm{d}_y f)(0,0)$ 是 $\mathbf{R}^n\to\mathbf{R}^n$ 的线性同构(即有逆映射存在，这里视 $x$ 为参数值)，则必存在 $0\in U$ 的一个连通邻域 $\Omega$. 以及唯一的光滑映射 $g:\Omega\to V$，使 $g(0)=0$，而且

$$f(x,g(x))=0. x\in\Omega. \tag{36}$$

**证** 考虑一个新映射

$$\varphi: U\times V\to\mathbf{R}^m\times\mathbf{R}^n$$
$$(x,y)\to(x,f(x,y)).$$

它在(0,0)处的雅可比矩阵是

$$J=\begin{pmatrix} I & 0 \\ * & (\mathrm{d}_y f)(0,0) \end{pmatrix}.$$

它自然是非退化的.于是由反函数存在定理知道 $\varphi^{-1}$ 在(0,0)附近存在:$\varphi^{-1}[(x,f(x,y))]=(x,y)$.由 $f(0,0)=0$,可知 $\varphi(0,0)=(0,0)$,今证 $\varphi^{-1}(x,0)$ 就是隐函数 $y=g(x)$ 的图(graph).一个函数 $y=g(x)$,$x\in U$,$y\in V$ 的图就是 $U\times V$ 的一个子集 $\{(x,y)\}\subset U\times V$,而其中不会含有两个形如 $(x,y_1)$,$(x,y_2)$ 的元 $(y_1\neq y_2)$.即是说同一个 $x$ 不会对应于两个不同的 $y_1$ 和 $y_2$.这正是函数定义之所以必须的.$\varphi^{-1}\{(x,0)\}$ 是隐函数图的证明如下:$\varphi^{-1}(x,0)$ 是一个 $m+n$ 维向量,它的前 $m$ 个分量就是 $(x,0)$ 的前 $m$ 个分量 $x$,若记后 $n$ 个分量为 $y$,则 $\varphi^{-1}(x,0)=(x,y)$.因此 $(x,y)$ 由 $(x,0)$ 单值地决定,从而 $y$ 由 $x$ 单值地决定,不会发生同一个 $x$ 对应于不同的 $y_1\neq y_2$ 的情况.所以 $y$ 是 $x$ 的某个函数 $g(x)$ 之值:$y=g(x)$.今证这就是方程(36)决定的隐函数.事实上,由 $\varphi(x,y)=(x,0)$,而 $\varphi$ 之定义给出 $\varphi(x,y)=f(x,y)$.所以

$$(x,0)=\varphi(x,y)=\varphi(x,g(x))=(x,f(x,g(x))),$$

即是说 $g(x)$ 适合(36)式,而且由 $\varphi$ 把(0,0)一对一地映为(0,0)知 $\varphi(0,g(0))=(0,0)$,即 $g(0)=0$.总之我们有 $\{\varphi^{-1}(x,0)\}\subset\{$隐函数 $g(x)$ 之图$\}$.

反过来还要证明{隐函数 $g(x)$ 之图$\}\subset\{\varphi^{-1}(x,0)\}$.事实上,若由(36)定义的隐函数存在,记为 $g(x)$,这里 $g(0)=0$.则由 $f(x,g(x))=0$ 有 $\varphi(x,g)=(x,f(x,g))=(x,0)$.所以 $(x,g)\in\{\varphi^{-1}(x,0)\}$,而上述包含关系得证.

总结以上所述可知由 $\varphi$ 之逆映射 $\varphi^{-1}$ 存在且为唯一即知(36)决定的且适合 $g(0)=0$ 的隐函数是存在与唯一的.现在由 $\varphi$ 及 $\varphi^{-1}$ 之可微性证明 $g(x)$ 在 $x=0$ 附近可微.由 $\varphi^{-1}$ 在(0,0)附近可微即知

$$\begin{aligned}(x+h,g(x+h))-(x,g(x))&=\varphi^{-1}(x+h,0)-\varphi^{-1}(x,0)\\&=(\mathrm{d}\varphi^{-1})(x,0)h+o(\|h\|).\end{aligned}$$

它们后 $n$ 个分量就给出

$$g(x+h)-g(x)=(\mathrm{d}_n\varphi^{-1})(0)h+o(\|h\|).$$

这里 $(\mathrm{d}_n\varphi^{-1})(0)$ 表示由 $(\mathrm{d}\varphi^{-1})(x,0)$ 之后 $n$ 行所成的矩阵所代表的线性映射.因此 $g$ 是可微的.

与定理 2 一样,当 $f\in C^k$ 或为复解析时,$g(x)$ 也是 $C^k$ 光滑或复解析的.

**5.牛顿方法**　上面我们提到了反函数问题与隐函数问题都涉及方程 $f(x)=y$ 或 $f(x,y)=0$ 的求解问题.定理 2 证明了一个基本存在定理,其基础是逐步逼近法.实际上,我们首先是把求解方程 $f(x)=y$ 的问题化为求映射 $x\to y-g(x)$ $(g(x)=f(x)-x)$ 的不动点,然后给出了一个迭代程序.

$$\begin{cases}\text{自 } B_{\frac{\varepsilon}{2}}(0)\text{ 中的任意 } x_0\text{ 开始,}\\ \text{令 } x_{k+1}=y-g(x_k)\end{cases}. \tag{37}$$

为了证明 $\{x_k\}$ 的收敛性,我们看到 $\mathrm{d}g(0)=0$ 从而 $\|(\mathrm{d}g)(x)\|$ 在 $B_{\frac{\varepsilon}{2}}(0)$ 中可以任意小(这里用的是 $\|g(x^*)-g(x^{**})\|\leqslant\frac{1}{2}\|x^*-x^{**}\|$),在这里起了决定作用.适合这种条件的映射称

为压缩映射，所以我们实际上证明了压缩映射有不动点存在．但是不动点问题是数学中的一个重大问题．压缩映射有不动点存在只是其最简单的一例．一个比较著名而且十分有用的是**布劳威尔不动点定理**：若 $f$ 是映 $\mathbf{R}^m$ 中的一个球到其自身的连续映射(不需要压缩性)，则 $f$ 必有至少一个不动点．但是由于它的本质与压缩映射定理完全不同，后者涉及度量空间的完备性，而前者则与连续映射的拓扑度有关，所以我们只把它作为第六章的一个附录来讲．

除此以外，压缩映射定理讲的是一个迭代程序(37)，而迭代程序有许多种，相应于求某一方程 $F(x)=0$ 的近似解的不同迭代方法．现在以一个自变量 $x$ 的一个函数 $F(x)=0$ 之解为例．设 $F(x)$ 是定义在 $[a,b]$ 上的连续函数，$F(a)\cdot F(b)<0$．微积分教科书中常讲到用割线法去求解 $F(x)=0$．我们可以取 $x_0=a$(若恰好 $F(x_0)=0$，则是一件很偶然的事)，然后连接 $(a,F(a))$，$(b,F(b))$，它必与 $y=0$ 有一个交点．设它是 $y=0$ 上之 $x_1$ 点．然后作 $(x_1,F(x_1))$ 又与 $(b,F(b))$ 相连，并设割线与 $y=0$ 交于 $x_2$ 点，仿此进行，得到一串近似的 $\{x_k\}$，$\{x_k\}$ 的极限 $\bar{x}$ 即是 $F(x)=0$ 之解(图 3-5-5)．

用公式写，过 $(a,F(a))$，$(b,F(b))$ 之割线为

$$y-F(a)=\frac{F(b)-F(a)}{b-a}(x-a).$$

所以

$$x_1=x_0-\frac{b-x_0}{F(b)-F(x_0)}F(x_0),$$

而一般的迭代式是

$$x_{k+1}=x_k-\frac{b-x_k}{F(b)-F(x_k)}F(x_k). \tag{38}$$

然后我们来证明 $\{x_k\}$ 有极限 $\bar{x}$ 存在，而 $\bar{x}$ 就是 $F(x)=0$ 的解．我们这里不来证明 $\bar{x}$ 的存在性，只是指出这个迭代程序的收敛速度是很慢的．注意，这个迭代与证明反函数存在时所用的迭代(37)是不一样的，而且也没有利用微分学的基本思想((37)中也没有利用微分，而只是在证明这个映射是压缩映射时才用到了它)，所以我们会问如何把微分学的基本思想用上去？注意到 $\frac{F(b)-F(x_k)}{b-x_k}\sim F'(\cdot)$(略去高阶无穷小量)，可否把这个式子代入(38)以得到一种新的迭代？问题是 $F'(\cdot)$ 中的"$\cdot$"是什么？一个方法是假设 $F'(\cdot)=F'(b)$，另一种是假设它为 $F'(x_k)$．这样就得到两种迭代程序．一种是最简单的．即

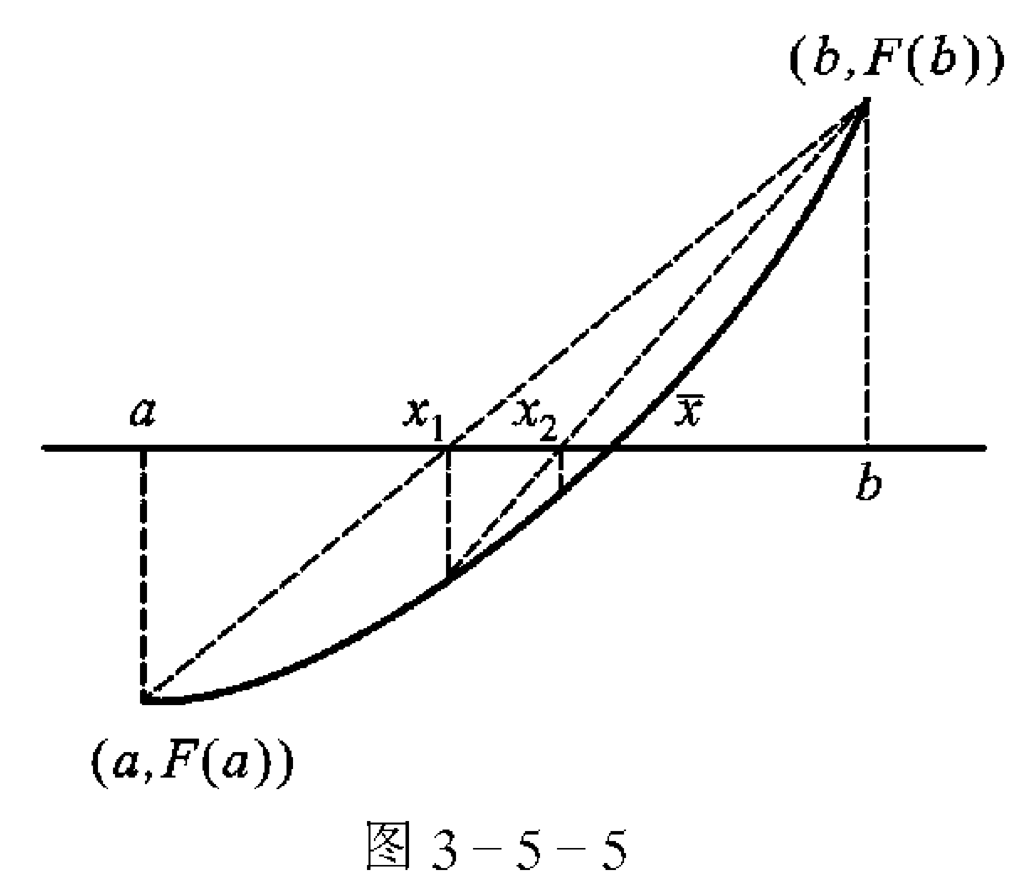


图 3-5-5

$$\begin{cases}\text{从 } x_0=b \text{ 开始}.\\ \text{令 } x_{k+1}=x_k+[F'(x_0)^{-1}(y-F(x_k))] \quad (y=0),\end{cases} \tag{39}$$

另一种即是著名的牛顿方法(Newton's mothod)：

$$\begin{cases}\text{从 } x_0=b \text{ 开始}.\\ \text{令 } x_{k+1}=x_k+[F'(x_k)]^{-1}(y-F(x_k)) \quad (y=0).\end{cases} \tag{40}$$

(39)则称为简化的牛顿方法.图 3－5－6 上说明了这两种迭代的几何意义.所以牛顿方法称为切线法.图上画的是函数 $F(x)=\frac{1}{4}(x^2-1)$的图形.$F(x)=0$ 的解是 $x=1$,如果用这两种迭代程序,而且都从 $x_0=2$ 开始,我们会得到

| 简化牛顿方法 | 牛顿方法 |
|---|---|
| $x_0=2$ | $x_0=2$ |
| $x_1=1.416\ 666\ 7$ | $x_1=1.416\ 666\ 67$ |
| $x_2=1.263\ 069\ 1$ | $x_2=1.110\ 534\ 41$ |
| $x_3=1.178\ 483\ 4$ | $x_3=1.010\ 636\ 768$ |
| $x_4=1.042\ 091\ 4$ | $x_4=1.000\ 111\ 557$ |
| $x_5=1.018\ 702\ 9$ | $x_5=1.000\ 000\ 012$ |

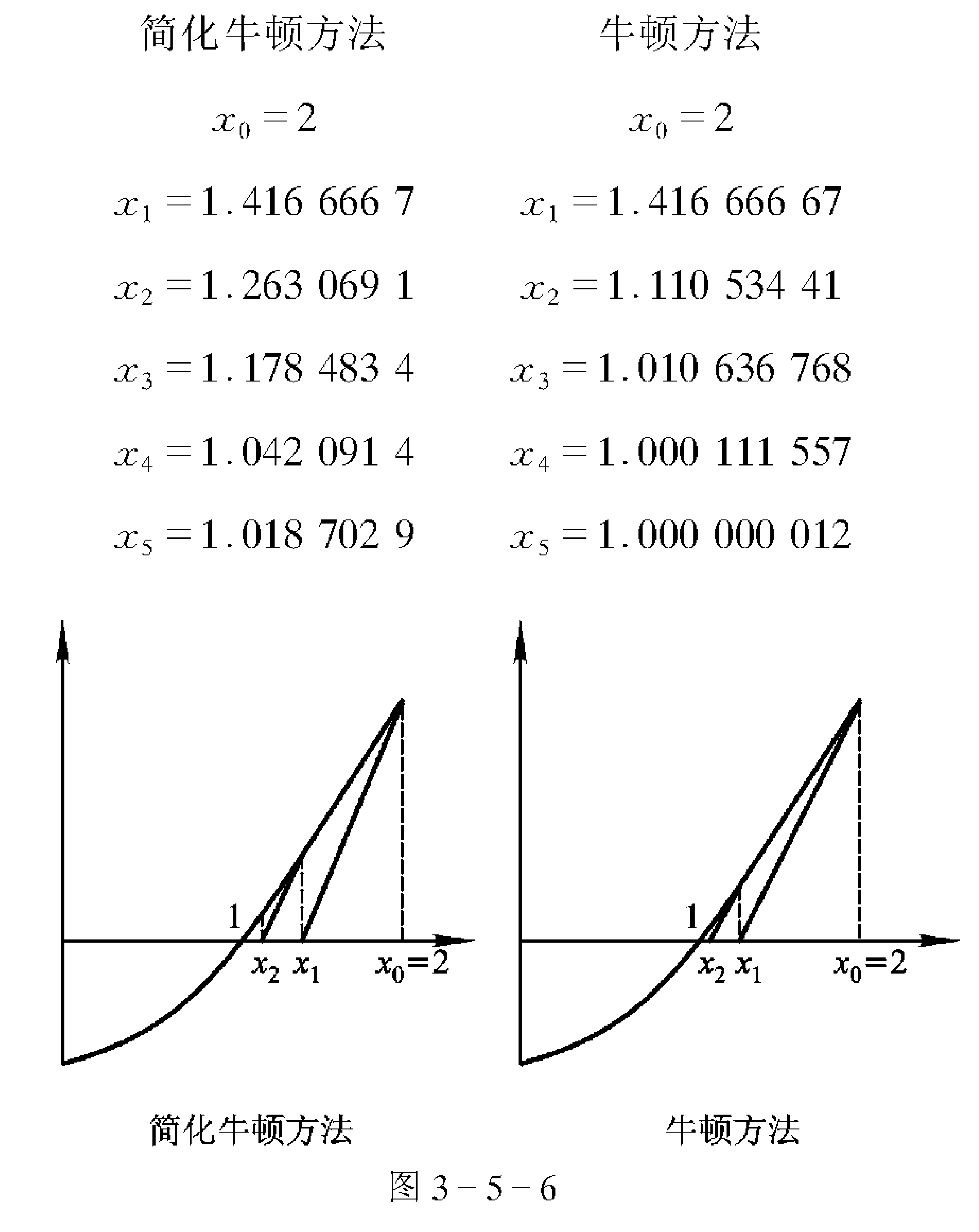



图 3－5－6

显然两种方法都收敛于 1.但是很明显,牛顿方法收敛得快多了.我们不来证明这里的结果,只指出:对于简化的牛顿方法我们有

$$|x_k-\bar{x}|\leqslant C\lambda^k, 0<\lambda<1.$$

而对于牛顿方法则有

$$|x_{k+1}-x_k|\leqslant \delta^{2^k-1}|x_1-x_0|^{2k}, \qquad 0<\delta,$$

$\delta$ 是一个适当的常数.因此,若$|x_1-x_0|$相当小,则$\{x_k\}$的收敛要快得多.

牛顿方法不但是一个数值计算方法,而且在讨论许多重要的自然界现象——例如太阳系的稳定性——时会给我们提供极重要的信息.整个说来,迭代方法引向了数学中一些广阔的领域——混沌、分形等等,这些当然都出了本书范围.

**6. 拉格朗日乘子法**　极值问题一直是微分学的重要问题,我们在前面即已讲过局部极值与最大最小值是有区别的:后者实际上是一个整体性的概念.但即令是局部概念,在最简单的光滑函数情况下,我们实际上是研究函数的临界点的问题.如果 $x_0$ 是 $f(x)$的临界点,$f(x)$在 $x_0$ 并不一定达到极值.所以有时我们称 $f(x)$在临界点处的值为平稳值(stationary value)而与极值

(extremum)相区别.

现在我们要考虑这样一个极值问题.设在 $x_0$ 的某邻域 $\Omega\subset\mathbf{R}^m$ 中定义一个函数 $f(x)$,可是我们并不是在 $x_0$ 的邻域中求 $f(x)$之极值或平稳值,而是在一些补充条件

$$\varphi_i(x)=0, i=1,2,\cdots,k \tag{41}$$

下求 $f(x)$之极值或平稳值.当然我们设 $x_0$ 适合条件(41).这在物理上是很自然的,条件(41)称为约束条件(constraints).如果 $k=1$,则条件(41)下求 $f(x)$之极值就是求 $f(x)$在曲面 $\varphi_1(x)=0$ 上的极值.这种极值问题称为条件极值问题.$\mathbf{R}^m$ 中的$(m-1)$维曲面称为一个超曲面.如果 $k>1$则(41)定义一个$(m-k)$维曲面(或者准确一些说是 $\mathbf{R}^m$ 的一个$(m-k)$维子流形).当然,(41)中的 $k$ 个条件应该是独立的.曲面、子流形等概念是以放在光滑函数的框架下处理最为方便,因为我们有完整的微分学的理论作为工具.因此,以下我们的目标函数 $f(x)$以及约束中的函数均设为光滑函数($C^k$ 甚至 $C^\infty$,这一点前面已经说过,但是现在不能讨论复值函数问题,因为复数不能比较大小,也就无所谓极值).上面说的"独立性"也是不清楚的.模仿前面关于单射定理与投影定理的讨论,我们把它定义为雅可比矩阵$\dfrac{\partial(\varphi_1,\cdots,\varphi_k)}{\partial(x_1,\cdots,x_m)}$在 $x_0$ 附近有最大可能的秩 $k$.这时自然会有 $k<m$.不可能 $k>m$,一方面是因为$\dfrac{\partial(\varphi_1,\cdots,\varphi_k)}{\partial(x_1,\cdots,x_m)}$矩阵之秩$\leqslant\min(m,k)$.同时,一个点 $x_0$ 只有 $m$ 个坐标.所以不可能满足多于 $m$ 个独立的方程.$k=m$ 也是没有意义的,因为在 $x_0$ 附近满足(41)中的 $m$ 个方程的点只有一个 $x_0$,这是反函数存在定理的简单推论.有了这样的约定后,我们就可以把(41)中的 $k$ 个函数看成一个映射

$$\Phi:\Omega\subset\mathbf{R}^m\to\mathbf{R}^k,(x_1,\cdots,x_m)\mapsto(\varphi_1,\cdots,\varphi_k), \tag{42}$$

于是适合(41)之点的集合就是 $\Phi^{-1}(0)$.在最大可能秩的条件下,它是 $\mathbf{R}^m$ 的一个$(m-k)$维子流形 $M_{m-k}$.且 $x_0\in M_{m-k}$.而所谓条件极值问题就是在$(m-k)$维子流形 $M_{m-k}$上之 $x_0$ 附近求 $f(x)$之极值或平稳值.

一个自然的想法是,既然映射(42)在 $x_0$ 附近有最大可能秩 $k$,我们可以假设雅可比矩阵的 $k$ 阶子矩阵

$$J=\begin{pmatrix}\dfrac{\partial\varphi_1}{\partial x_1} & \cdots & \dfrac{\partial\varphi_1}{\partial x_k}\\ \\ \dfrac{\partial\varphi_k}{\partial x_1} & \cdots & \dfrac{\partial\varphi_k}{\partial x_k}\end{pmatrix}$$

在 $x_0$ 处非退化,从而由隐函数定理,由(41)可以解出

$$x_i=\Psi_i(x_{k+1},\cdots,x_m)\qquad i=1,2,\cdots,k \tag{43}$$

适合

$$x_i^0=\Psi_i(x_{k+1}^0,\cdots,x_m^0).$$

$(x_1^0,\cdots,x_m^0)$是 $x_0$ 的 $m$ 个坐标.以(43)代入目标函数,并记 $x'=(x_{k+1},\cdots,x_m)\in\mathbf{R}^{m-k}$,我们会得到定义于 $x'=x_0'\in\mathbf{R}^{m-k}$附近的新目标函数

$$F(x')=f[\Psi_1(x'),\cdots,\Psi_k(x'),x']. \tag{44}$$

于是条件极值问题就化成了通常的极值问题.

但是我们总是想避免应用隐函数定理,因为把$(x_1,\cdots,x_k)$解出成为$x'$的函数(43)是太困难了.好在我们有

**定理 6** 设$\Omega\subset\mathbf{R}^m$是0的一个充分小的邻域.$f$在$\Omega$中光滑.又设有(42)式定义的由$\Omega$到$\mathbf{R}^k$的光滑映射$\Phi$,$\Phi(0)=0$,且在$\Omega$中适合最大秩条件.若$f$在约束$\varphi_i=0$下在$x=0$处取局部极值(平稳值),则必存在$k$个常数$(\lambda_1,\cdots,\lambda_k)$,使得

$$\mathrm{d}\left[f-\sum_{i=1}^{k}\lambda_i\varphi_i\right](0)=0. \tag{45}$$

$(\lambda_1,\cdots,\lambda_k)$称为拉格朗日乘子(Lagrange multipliers).

**证** 为计算简单起见,令$m=2,k=1$.于是目标函数$f=f(x,y)$.约束条件只有一个$\varphi(x,y)=0$.最大秩条件是$(\varphi_x,\varphi_y)\neq 0$.我们不妨设$\varphi_x(x,y)\neq 0$于$(0,0)$附近.于是由约束条件可以解出$x=\psi(y)$,$0=\psi(0)$,而$x'=y$.目标函数变成

$$F(y)=f[\psi(y),y].$$

现在$y=0$变成了$F(y)$的极值点,所以

$$F'(0)=\frac{\partial f}{\partial y}(0,0)+\frac{\partial f}{\partial x}(0,0)\psi'(0)=0. \tag{46}$$

但由约束条件,在$y=0$附近

$$\varphi[\psi(y),y]=0,$$

从而

$$\varphi_y(0,0)+\varphi_x(0,0)\psi'(0)=0.$$

最大秩条件告诉我们

$$\psi'(0)=-\varphi_y(0,0)/\varphi_x(0,0).$$

代入(46),并令

$$\lambda=f_x(0,0)/\varphi_x(0,0) \tag{47}$$

即得

$$f_y(0,0)-\lambda\varphi_y(0,0)=0.$$

(47)本身就是

$$f_x(0,0)-\lambda\varphi_x(0,0)=0.$$

两式合在一起即(45)式.定理证毕.

拉格朗日乘子之所以重要不仅在于它使我们不必去构造隐函数$\Psi_1(x'),\cdots,\Psi_k(x')$,而只需要从(45)(其中共有$m$个方程)以及约束条件$\varphi_i(x)=0$(其中$i=1,2,\cdots,k$)求出$m+k$个未知数$(x_1^0,\cdots,x_m^0,;\lambda_1,\cdots,\lambda_k)$(这些数同样不容易求);而在于拉格朗日乘子时常有重要的物理和几何意义.下面看一个重要例子:关于二次曲面的极值性质——这些性质在物理上是很有用的.

设有$\mathbf{R}^m$中的有心二次曲面

$$Q:\sum_{i,j=1}^{m}a_{ij}x_ix_j=1,\qquad a_{ij}=a_{ji}. \tag{48}$$

所谓"有心",即$\det[a_{ij}]\neq 0$,这样把有抛物性质的退化情况全排除在外了.(48)又可用内积形式

写成

$$Q: \langle Ax, x\rangle = 1,\ A = (a_{ij}).\tag{48'}$$

现在在 $Q$ 上求一点使 $\|x\|^2 = \langle x, x\rangle = \sum_{i=1}^{m} x_i^2$ 取平稳值例如极大.这就是一个条件极值问题,目标函数是 $f(x) = \|x\|^2$,(48)就是约束条件,所以 $k=1$.如果直接应用拉格朗日乘子方法,就应考虑

$$F(x) = \|x\|^2 - \lambda\langle Ax, x\rangle.$$

为了说明拉格朗日乘子的意义我们可以把它化为一个对偶问题.为此令

$$G(x) = \frac{\langle Ax, x\rangle}{\|x\|^2},$$

而原来的问题:在 $\langle Ax, x\rangle = 1$ 上求一点 $x$,使 $\|x\|^2 \to$ 极大,现在变成设法去找 $x_0$ 使

$$G(x_0) \to \min.$$

如果记 $y = x/\|x\|$,则必有 $\|y\|^2 = 1$,而 $G(x)$ 变成

$$H(y) = \langle Ay, y\rangle.$$

易见 $H(y) = G(x)$.求 $G(x)$ 之极小值,就化为在约束条件 $\|y\|^2 = 1$ 下,求 $H(y)$ 的极小值.关于 $H$ 的极值问题称为原问题的对偶问题,于是我们对

$$\tilde{F}(y) = \langle Ay, y\rangle - \mu\|y\|^2$$

应用拉格朗日乘子法,而有

$$\frac{\partial \tilde{F}}{\partial y_i} = 0 \quad \text{即} \quad \sum_{j=1}^{m} a_{ij}y_j - \mu y_i = 0,\ i = 1, 2, \cdots, m.$$

用矩阵表示,即

$$(A - \mu I)y = 0.\tag{49}$$

然后再求 $m+1$ 个未知数 $(y_1, \cdots, y_m, \mu)$.由线性代数知识可见,上式之解中 $\mu$ 是 $A$ 的特征根,$y$ 是相应的范数为 1 的特征向量.回到 $F$,则拉格朗日乘子 $\lambda$ 相当于 $\frac{1}{\mu}$.而 $x$ 仍表示同一个特征向量,不过 $\|x\|^2$ 不再是 1.现在 $A$ 是对称矩阵,它一定有 $m$ 个实特征根,依大小排列为

$$\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_m.$$

我们用以上方法求出的拉格朗日乘子就是这些 $\mu$ 和相应的 $\lambda$.如果已求出拉格朗日乘子 $\lambda_1 = \frac{1}{\mu_1}$,相应的 $x$ 和 $y$ 记为 $x^{(1)}$ 和 $y^{(1)}$,约束条件成为 $\langle Ax^{(1)}, x^{(1)}\rangle = 1$ 和 $\|y^{(1)}\|^2 = 1$.$x^{(1)}$ 和 $y^{(1)}$ 是相应于相同特征根的特征向量.$x^{(1)}$ 适合 $\langle Ax^{(1)}, x^{(1)}\rangle = 1$ 就是说向量 $x^{(1)}$ 的两个端点一个在是 0,另一端在二次曲面 $\langle Ax, x\rangle = 1$ 上.这样一来,原来的条件极值问题就成为在此二次曲面上求一点.使由原点(即二次曲面的中心)到此点之距离为极大.以上所得的结果就是:所求的 $x$ 是一个特征向量,而利用 $G(x_0) \to \min$ 知这个距离的平方即特征根的倒数,这样求出的特征向量称为二次曲面的主半轴.但是有一些特征根为负,就表示有一些主半轴之长的平方为负,这种主半轴就称为虚轴.

以上我们只求出了一个主轴,但实际上有 $m$ 个主轴.其它主轴可化为再附加另一些约束条件的条件极值问题,我们就不在这里讨论了.

# §6　变分法大意

**1. 变分原理**　本章一开始就举了两个古典的例子，它们不但是微分学历史上两个重大例子，确切地体现了微分学的基本思想——舍去高阶无穷小，而且还有更深刻的含意：它们表现了两种不同的方法论．第一个例子是牛顿对面积速度不变原理（即角动量守恒）的证明．牛顿把时间分成一小段一小段．第一个小段结束瞬间的运动状况（即质点的位置和速度），即下一小段时间起始时的初始状况．如此推演下去即可穷尽这个质点运动的历史．我们一向认为这是机械论的决定论的典型表现．当时，科学的发展也只能使人们得到这样的结论．但是实际情况并不如此简单．例如我们在§1中讨论开普勒三定律时，只讨论了一个太阳对一个行星的作用．实际上其它行星、卫星的影响都忽略不计了．开普勒主要是从对火星的观测中得到的他的三个定律，如果他观测的是金星和水星，则比较容易发现天体的实际运动与他所主张的椭圆有不可忽略的差距．拉普拉斯与拉格朗日都知道，理论上的椭圆轨道会受到“微扰”（perturbation）．到了19世纪末，庞加莱开始研究：微扰的长期积累的后果会是什么？他的研究汇集成他的名著《天体力学的新方法》，开辟了数学的新篇章．其实后来人们研究得很多的“混沌”现象等等在这里都已出现．在此以前，牛顿力学意味着宇宙的一切从初始时间，即上帝的手作了“第一次推动”（the first blow）以后，就一切都决定了，再无任何变化．所以说“阳光之下没有新事物”（这句话源出圣经）．用数学的语言表述，这种“机械论的决定论”其实就是微分方程的柯西问题

$$\frac{\mathrm{d}x}{\mathrm{d}t}=f(t,x),$$

$$x|_{t=0}=x_0$$

的解的存在性与唯一性．到了19世纪末，就提出了新问题：这些解确实可以一直到$t\to+\infty$时都存在，还是在某一时刻$t\to T$就会发现“爆破”（blow up）？解当$t\to+\infty$时的渐近状况如何？全局状况如何？“机械论的决定论”一定意味着可以精确地计算解吗？随机性是否以某种方式存在于决定性之中，或者二者以某种方式相联系？总之，现时人们对“机械的决定论”的习惯的理解中，这些问题都没有提出来．但是自牛顿以来的这种方法论今天的内涵自非当年之可比．可惜的是，本书不可能再对此稍作介绍了．

第二个例子却是另一种方法论的典型问题．当费马研究光由$A$到$B$的折射时，他遵循了另一种方法论．当光才离开$A$点后，下一个时刻$t$光的位置在哪里？我们现在知道，光的“位置”一说很不清楚，只能说光的波前是以$A$为心，$ct$为半径（$c$是光速）的球面．再往后，则光的波前上的每一点都成了次级光源，于是再经过时间$\tau$这些次级光源又各产生自己的波前：以次级光源为心，$c\tau$为半径的球面，把这些新波前的包络找出来，得到一个以$A$为心，$c(t+\tau)$为半径的球面，是为光在时刻$t+\tau$的波前．大家都知道，这就是惠更斯原理．它的出发点是：光是一种波，用这个方法也可以讨论光的直线行进，光的折射反射，但完全是用几何方法，主要是用包络．但是，关于光的本性还有另一种学说即粒子说，它认为光的本性是粒子．牛顿虽然是粒子说的拥护者，但是没有见到谁把牛顿的运动定律用于光的粒子的解说．代替它，费马提出了关于光的传播的最短时间原理（即费马原理）：“光在一切可能的路径中必取耗时最短的路线”．所谓可能的路径即

不违背所设问题中的给定约束的路径. 在折射问题中即由 $A$ 到折射面上一点 $L$，这一段路径是在某种介质(例如空气)中，而在其中光走直线，然后再由 $L$ 在另一种介质(例如水)中走到 $B$. 因此又走另一段直线. 现在的问题是怎样从可能的路径中找到实际发生的路径？用费马原理. 本章§1 中详细解释了这是怎样做到的，介绍了微分学的基本思想是怎样在起作用.

现在回想一下，费马原理实在是一个聪明的办法：我们不必去问，光究竟是不是粒子，质量是多少，受什么力的作用，初始位置与初速(包括方向)是多少？如果要这样做该有多少困难？讲反射还比较好说，不妨把光的反射解释为完全弹性的球在刚性边界上的反弹. 折射又如何呢？这必定会涉及介质的界面的作用. 现在好了，我们完全不必去管这些事：只要能节省时间就行！作为波动的光有此性质，作为粒子的例如天体是否也适合类似的原理呢？至少牛顿第一定律可作如是观：如果粒子在一均匀介质中运动而不受外力影响，它也一定走耗时最短的路径！因此人们想到，大自然的根本规律是否也是一个变分原理？而且把这归之于例如上帝最节约时间等等. 但是17—18 世纪终究是科学已开始摆脱玄学和宗教束缚的时代，人们再也不会满足于用抽象的思辨来代替对具体问题的研究了. 于是出现了法国人莫培督(Pierre - Louis Moreau de Maupertuis，1698—1759). 他在 1746 年的一篇论文“由形而上学原理导出的若干运动和静止的法则”中提出了一个他认为放诸整个宇宙而皆准的“一般原理”如下：“如果在自然界发生了某一变化，则此变化所需的作用量(action)必为尽可能的小.” 他并且认为这个原理正是上帝存在的证据. 但是，莫培督生活的年代已是牛顿力学风靡一时的年代，再只用宗教与玄学来回避问题已经不可能了. 何况莫培督本人也是科学圈中人. 他曾经率队进入北极圈，以实际数据来证实牛顿关于地球在两极比较扁平的论断. 他曾任普鲁士科学院主席，是欧拉的顶头上司，而且把欧拉关于变分学的研究成果说是应用了他的最小作用原理. 后来又陷入了一场大争论，终于 1759 年心力交瘁而病逝于约翰二世伯努利家中. (这里附带发一点议论，许多“知识产权”之争现在回想起来确实没有意义. 牛顿和莱布尼茨之争是如此，但是又有谁想到，今天我们称为牛顿运动方程的那一组方程其实是欧拉而不是牛顿本人写出来的？这是特鲁斯德尔研究的结论，见《天遇》一书(迪亚库等著、王兰宇译，上海科技教育出版社，2001，230 页)；哈密顿系统归之于拉格朗日更恰当. 诸如此类的事在数学历史上比比皆是. 我们只能得出这样的结论：重要的是科学，科学家的名声只是第二位的).

莫培督的最小作用原理真正的缺点是它不能解决很多问题. 他定义“作用”是这样的物理量：作用 = 质量 × 速度 × 距离. 试想如果扛着重物——很大的 $M$，很快地——很大的 $V$，走了很长的路——很大的 $L$，其作用自然很大. 从量纲来说，其量纲是 $[M][L][T]^{-1}[L]=[M][\frac{L}{T}]^2[T]$ 因此本质上是动能 × 时间. 如果再考虑作用了多长时间，则作用量 $A=\int_{t_0}^{T}\frac{m}{2}V^2\mathrm{d}t$. 但是无论是按费马原理还是最小作用原理时都会得到错误结论. 例如达西(Chevalier d' Arcy)就指出，莫培督的例子中有错. 例如有一个球面镜，球心为 $A$. 若在球心处置一光源，而让光从球的凹面反射. 费马原理或最小作用原理得出的规律应是入射角 = 反射角. 如此，则由 $A$ 到 $B$ 的反射点应是 $AB$ 联线与球面的交点 $Q$. 这样 $QAB$ 三点共线，成一半径的延长线，而光走过的路径是

$$AQ+QA+AB=2R+AB$$

(图 3-6-1 上). 但是若在球面上取任一点 $P$，则 $PB\leqslant PA+AB$，因此现在光走的路径是 $PA+PB<2R+AB$. 所以，费马原理给出的不但不是耗时最少，反而是耗时最多. 但是若 $A$ 在镜外(图

3－6－1下)，则由入射角＝反射角确实给出耗时最短的路径.于是达西就问道，为什么上帝遇到凹的镜面就浪费时间，而在凸的镜面上就节约时间.这个问题我们今天看来就很清楚，实际上的答案并非极小值而是平稳值.但在当时，这个批驳确实是很有力的.因此，最小作用原理确实需要认真地修改.它的基本思想是如此有力与新奇，所以许多第一流的数学家特别是拉格朗日做了许多工作.在他的名著《解析力学》一书中对最小作用原理有系统的论述.至此，我们看到一个变分原理(即某个量——作用量达到极大或极小)成了一个科学分支——现在是牛顿力学——的基础.这是有重大意义的.

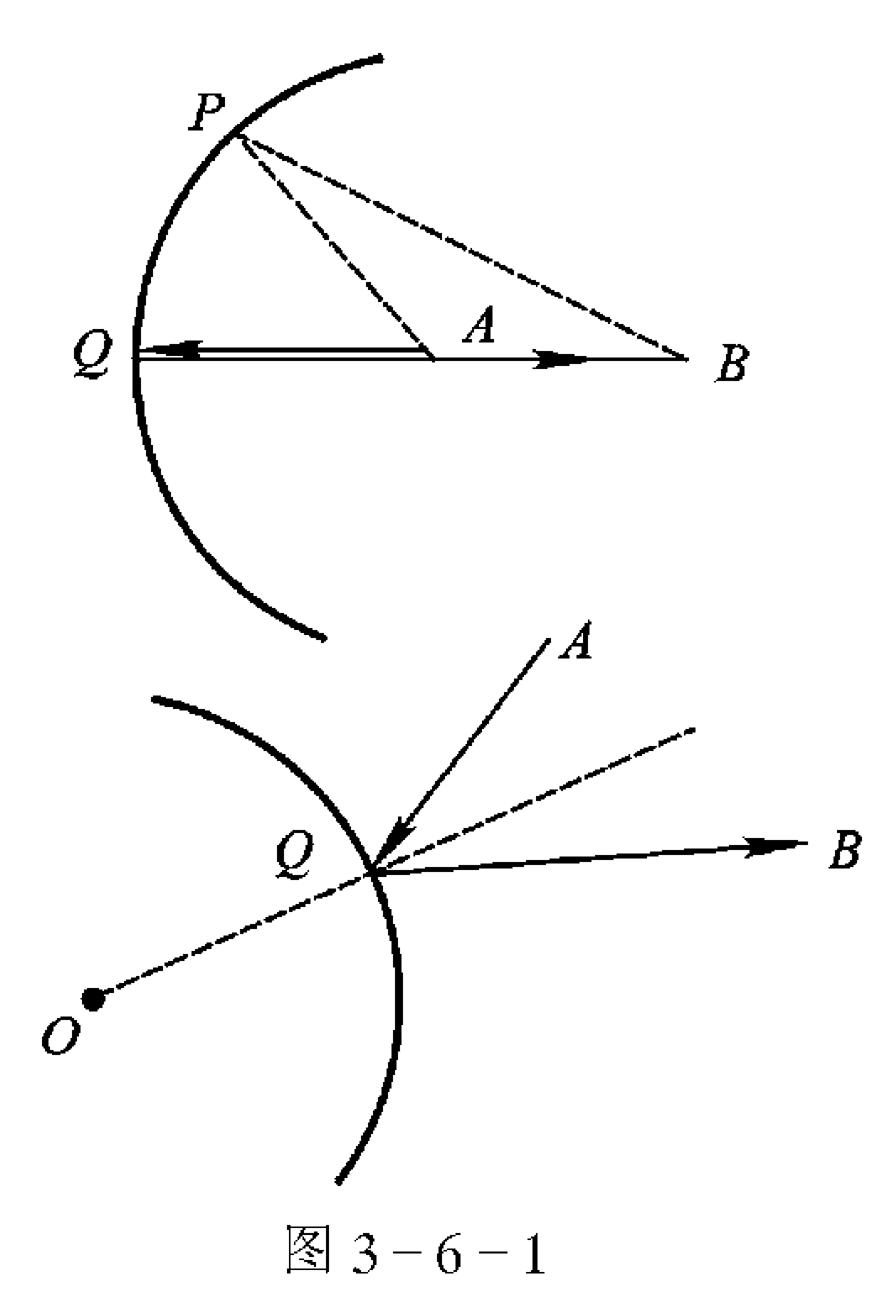


图3－6－1

**2．欧拉-拉格朗日方程** 古典的变分法(或称变分学)讨论的是一类可以写成积分形式的泛函之极值(平稳值)问题.下面说的捷线问题(brachistochrone)是其中著名的一个.它有很长的历史，伽利略研究过它，但是答案是错误的.下面讲的是约翰•伯努利一世(上面说到的莫培督病逝于其家中的是约翰二世.是约翰一世的儿子)对这个问题的解决，其方法与费马解决光的折射定律的方法多有相似而对后世很有影响.所谓捷线问题是这样的.设在一铅直平面上有 $A$，$B$ 两点，但 $B$ 不在 $A$ 之正下方，求一曲线 $L$ 使一质点 $P$ 在重力作用下沿 $L$ 从 $A$ 下降到 $B$ 所需时间最少(这里空气阻力与摩擦力均忽略不计).伽利略研究这个问题时指出，人们可能以为 $L$ 应该是联结 $A$，$B$ 两点的弦.他说这是不对的，$L$ 不是弦而是圆弧.伽利略前一句话是对的，后一句却错了.为什么不是弦呢？因为下降过程中在 $y$ 方向有加速度(见图3－6－2)，越降越快，如果我们用水平的直线 $y=y_i$，$i=0,1,\cdots,N$，划分平面，则在 $A$ 附近，因为下降速度慢，曲线应该陡一些，以便在较短时间内下降更多的高度.在 $B$ 点附近，质点速度已经很快，$L$ 平一些也可以很快到达终点.总之，$L$ 不是直线.为什么后一句错了？我们应该具体计算来看结果是什么.设所求曲线之方程为 $y=f(x)$，质点 $P$ 之速度为 $v$，质量为 $m$，则当 $P$ 到达 $L$ 上一点

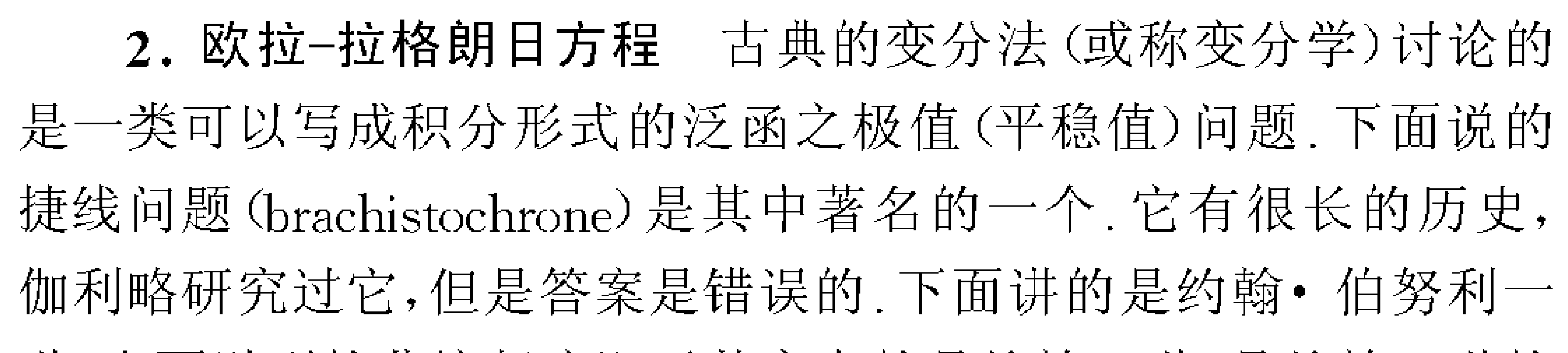


图3－6－2

$(x,f(x))$时，动能由 $A$ 点处 $P$ 静止时的0增到$\frac{1}{2}mv^2$，位能则由 $A$ 处的0减少到 $-mgf(x)$.所以

$$\frac{1}{2}mv^2=mgf(x),\quad v=\sqrt{2gf(x)}.$$

伯努利把 $L$ 分成许多小段所成的折线，而折线之顶点为 $(x_i,y_i)=(x_i,f(x_i))$，$i=0,1,\cdots,N$.$i=0$ 和 $N$ 相当于 $A$，$B$ 两点.第 $i$ 小段之长为$\sqrt{(x_i-x_{i-1})^2+(y_i-y_{i-1})^2}=\sqrt{(x_i-x_{i-1})^2+[f(x_i)-f(x_{i-1})]^2}$.不妨设在这一小段中速度不变而为$\sqrt{2gf(x_i)}$，于是这一小段耗时$\sqrt{(x_i-x_{i-1})^2+[f(x_i)-f(x_{i-1})]^2}/\sqrt{2gf(x_i)}$，而由 $A$ 到 $B$ 所需的时间是

$$T=\sum_{i=1}^{N}\sqrt{(y_i-y_{i-1})^2+(x_i-x_{i-1})^2}/\sqrt{2gy_i}.$$

因为 $y_i = f(x_i)$，所以 $T$ 是 $(x_1, x_2, \cdots, x_{N-1})$ 的函数，而问题就是求 $f$，使 $T \to \min$. 本来是求未知函数 $f(x)$，现在化成了求 $\mathbf{R}^{N-1}$ 空间中的一个点 $(x_1, x_2, \cdots, x_{N-1})$. 令 $N \to \infty$，就知道，我们的问题是求∞维空间的一"点"(即一个函数)，使 $T \to \min$. 但当 $N \to \infty$ 时

$$\sqrt{(y_i - y_{i-1})^2 + (x_i - x_{i-1})^2} \approx \sqrt{1 + [f'(x_i)]^2} \mathrm{d}x$$

所以 $T$ 趋向一个"泛函"

$$J = \int_0^{x_B} \sqrt{1 + [f'(x)^2]} \mathrm{d}x / \sqrt{2gf(x)}. \tag{1}$$

伯努利解决了这个问题，指出 $L$ 应该是旋轮线而不是圆弧，所以伽利略的后一句话错了.

伯努利的解法下面再说. 现在把它与 §1 中费马的折射定律比较一下：费马说的是一切可能的曲线即由 $A$ 经过折射面到达 $B$ 的折线，现在则是通过 $A$，$B$ 两点的曲线 $L$. 费马说的是光耗时最短的路径，现在则是使 $J \to \min$ 的曲线. 总之，这里有两个要素：一是一个可容许函数集 $M$，二是一个"目标"，现在是一个泛函. 而我们的问题应表述为：在 $M$ 中求一个元，使 $J \to$ 平稳值(不一定是极值). 所谓变分学的古典问题是 $J$ 可以表示成一个积分

$$J = \int_a^b F(t, u(t), \dot{u}(t)) \mathrm{d}t. \tag{2}$$

这里有一个重要的说明：$u$ 的光滑性与 $F$ 的光滑性极为重要，而且带来了数学分析中极细致的问题. 下面我们只讨论这个影响整个数学发展的问题的基本思想，关于其技术细节方面则把条件放得很宽. 请注意，读者可能需要从其它文献中找到所需的细微的处理，而这些处理时常是很难的. 现在则应把注意力放在基本思想上.

于是设有 $\mathbf{R}^d$ 中曲线 $u(t) = (u_1(t), \cdots, u_d(t))$. 设 $u \in C^3([a, b])$，$F(t, u, p) \in C^2([a, b] \times \mathbf{R}^d \times \mathbf{R}^d)$. 如果 $p = \dot{u}(t) = (\dot{u}_1(t), \cdots, \dot{u}_d(t))$ 则 $F(t, u(t), \dot{u}(t))$ 对 $t$ 为二阶连续可微. 记 $M = \{u(t) = (u_1(t), \cdots, u_d(t)) \in C^1([a, b]), u(a) = A, u(b) = B\}$，即 $M$ 是经过 $(a, A)$，$(b, B)$ 两点的 $C^1$ 曲线之集合，这里 $a \leqslant t \leqslant b$ 是这些曲线的参数的变化域. $M$ 称为可容许函数集. 在 $M$ 上积分(2)是有意义的，所以(2)是 $M$ 上的一个泛函. 现在的问题是求 $M$ 中一个元 $u_0$，使 $J(u_0) = \min$(或平稳值)，准确些说是要求 $J(u_0)$ 在 $u_0$"附近"(粗略地说，即适合 $|\dot{u}(t) - \dot{u}(t_0)|$ 及 $|u(t) - u(t_0)|$ 都对 $t$ 一致地很小的函数 $u(t)$ 之集合)是 $J(u)$ 的极小值，即局部极小值. 这就是古典的一维变分问题. 一维，这里说一维是指 $M$ 中之元是一维曲线，与上面所说变分问题是无穷维空间中的极值问题并不矛盾. 所以也可以考虑 $n$ 维变分问题. 这时(2)中 $F = (x, u, p)$，$x \in \Omega \subset \mathbf{R}^n$，$u \in \mathbf{R}^d$，而 $p$ 是一个 $d$ 维向量的各分量之一阶偏导数所成的向量，所以 $p \in \mathbf{R}^{nd}$. 而 $F$ 在 $\Omega \times \mathbf{R}^d \times \mathbf{R}^{nd}$ 中具有二阶连续偏导数.

$M$ 中的附加条件

$$u(a) = A, u(b) = B \tag{3}$$

是一组边值条件. 在 $n$ 维变分问题中，这个条件要代以 $u$ 在 $\partial\Omega$ 上适合的某种条件，例如

$$u|_{\partial\Omega} = g, g \text{ 是定义在} \partial\Omega \text{ 上的已知函数}. \tag{4}$$

(3)和(4)都是所谓的狄利克雷条件. $M$ 中的边值条件可以是多种多样的. 甚至有时对 $M$ 中的元在边界($t = a$，$t = b$ 或 $\partial\Omega$)上可以不加任何条件. 这就是所谓"自由边值条件". 要注意 $M$ 一般地说并不是线性空间. 因为边值条件(3)或(4)一般都是非齐次的，因此 $u$ 中任意二元之和一般不

再适合边值条件.但是 $M$ 一定是某线性空间的子集.例如按我们上面的陈述方法,应该取 $M\subset C^k([a,b])$,而对 $u\in C^1([a,b])$,我们通常定义其"大小"为

$$\| u \|_{C^1} = \max_{[a,b]} \| u(t) \| + \max_{[a,b]} \| \dot{u}(t) \| .$$

右方的 $\| u(t) \|$ 是指一个 $d$ 维向量的范数.这个线性空间的选择对变分问题的处理是十分重要的.

当然 $F$ 中也可以含有 $u$ 之高阶偏导数.这对下面的讨论影响不大,所以我们以下只限于讨论最简单的情况.现在 $J$ 是由 $M$ 到 $\mathbf{R}$ 的映射(我们限于 $t$ 及 $F$ 取实值的情况,因为复数不能比较大小,所以谈不上极值. $u,p$ 等均设为实向量,不然的话就分开它们的实虚都来处理好了):

$$J: M\to \mathbf{R}.$$

正如在微分学中讨论函数的极值时,我们取一阶微分而舍去高阶无穷小量,现在我们也考虑

$$J(u+\delta u)-J(u).$$

这里的 $\delta u$ 就是以前说过的 $h$,把它写成 $\delta u$ 使人想起 $\mathrm{d}x$ 以及随之而来的种种争论.现在很明白了,它就是适合 $\| \delta u \|_{C^1}$ 可以任意小的函数,也就是上文所谓 $u_0$ 的"附近"(还有边值条件的考虑见下文).在老的书上有不少关于 $\delta u$ 的论述,例如 $\mathrm{d}\delta u=\delta \mathrm{d} u$ 等等,现在看来都毫无必要了.但我们仍然使用这个记号,而且称之为变分(variation).其实就是在另一个空间例如 $C^1$ 中的 $h$, $\delta u$ 虽然是历史的"遗迹",但它提醒我们,将来会考虑它趋于 0 时的情况,而一切有用的结果尽在极限中.实际上,以后我们总是令

$$\delta u=\lambda\varphi, \tag{5}$$

这里 $\varphi\in C^1$ 而且适合"齐次"边值条件(现在的情况即 $\varphi=\varphi(t)$, $\varphi(a)=\varphi(b)=0$).以后只需令 $\lambda\to 0$ 即可.

在作了限制(5)以后,$J(u+\delta u)$就变成了 $\lambda$ 的函数

$$\Psi(\lambda)=J(u+\lambda\varphi)=\int_a^b F(t,u+\lambda\varphi,\dot{u}+\lambda\dot{\varphi})\mathrm{d}t. \tag{6}$$

所以 $J(u)\to\min$,就变成简单的微分学问题:$\Psi'(0)=0$.但

$$\Psi'(\lambda)\big|_{\lambda=0}=\int_a^b [F_u(t,u,\dot{u})\varphi+F_p(t,u,\dot{u})\dot{\varphi}]\mathrm{d}t$$

对后一项作分部积分法,注意到 $\varphi\in C^1$ 且 $\varphi(a)=\varphi(b)=0$,有

$$\begin{aligned}\Psi'(0)&=\int_a^b [F_u(t,u,\dot{u})\varphi-\frac{\mathrm{d}}{\mathrm{d}t}F_p(t,u,\dot{u})\varphi]\mathrm{d}t+F_p(t,u,\dot{u})\varphi\Big|_a^b\\&=\int_a^b (F_u-\frac{\mathrm{d}}{\mathrm{d}t}F_p)\varphi\mathrm{d}t=0.\end{aligned} \tag{7}$$

现在要解释一下为什么对 $F$ 及 $u$ 之光滑性作如上的假设.首先是 $F$ 对其一切变元属于 $C^2$,这样保证了例如$\frac{\mathrm{d}}{\mathrm{d}t}F_p$ 有意义.但是$\frac{\mathrm{d}}{\mathrm{d}t}F_p$ 中含有对 $u$ 的二阶导数,而我们只假设了 $M$ 中之元 $u\in C^1$.为什么不在一开始就规定 $M$ 中的元 $u\in C^2$ 呢?因为这样 $M$ 就太小,而在 $M$ 中找我们需要的解就难多了.所以数学中为了解决存在性问题,考虑的范围以大一点为宜,而不能搞"武大郎开店"的办法.反过来,要解决唯一性问题时,武大郎的办法就起作用了.症结在于适当选 $M$.这就是为什么本节中的 $M$ 中之元 $u\in C^1$.但这样一来就有矛盾,这个矛盾的解决即所谓正则性问题,

下面还要说，这些情况在讲到狄利克雷原理时会看得更清楚.

再往下我们证明一个重要的引理(前面我们说过，数学中似乎有一个“规矩”，大凡重要的结果总是称为“引理”的，这里又是一例).

**引理 1**(**变分学的基本引理. 杜•波瓦•雷蒙** *Du Bois-Reymond*) 设 $H(x)$ 是定义于 $[a,b]$ 上的连续 $d$ 维向量值函数：$H:[a,b]\to\mathbf{R}^d$. 若对一切适合 $\varphi(a)=\varphi(b)=0$ 的连续 $d$ 维向量函数 $\varphi(x)$ 均有

$$\int_a^b H(x)\varphi(x)\mathrm{d}x=0,$$

则必有 $H(x)\equiv 0$ 于 $[a,b]$ 上.

**证** 用反证法. 设 $H(x)$ 有一个分量 $H_1(x)$ 在 $x_0\in(a,b)$ 处为正，则必可选 $\delta>0$ 充分小，使 $(x_0-\delta,x_0+\delta)\in[a,b]$，而且 $H_1(x)$ 在 $(x_0-\delta,x_0+\delta)$ 中仍为正. 今取 $\varphi(x)=(\varphi_1(x),\cdots,\varphi_d(x))$ 如下：

$$\varphi_1(x)=\begin{cases}0, & |x-x_0|\geqslant\delta,\\ \text{正}, & |x-x_0|<\delta,\end{cases}$$

$$\varphi_i(x)=0, i=2,\cdots,d.$$

于是一方面由假设

$$\int_a^b H(x)\varphi(x)\mathrm{d}x=0,$$

另一方面

$$\int_a^b H(x)\varphi(x)\mathrm{d}x=\int_{x_0+\delta}^{x_0+\delta} H_1(x)\varphi_1(x)\mathrm{d}x>0.$$

这个矛盾证明了引理.

把这个引理用于(7)式即有

**定理 2**(**极值函数的必要条件**) 设 $F(t,u,p)$ 及 $M$ 已如上述. 若 $u(t)$ 使 $J(u)\to\min.$ 则必有

$$F_u(t,u,\dot{u})-\frac{\mathrm{d}}{\mathrm{d}t}F_p(t,u,\dot{u})=0. \tag{8}$$

(8) 式是一个二阶常微分方程组，称为上述变分问题的欧拉—拉格朗日方程组(Euler-Lagrange，以下简记为 E－L 方程组).

上面已经提出了一个问题，即 $M$ 中的元 $u\in C^1$，则它怎么可能满足二阶的 E—L 方程组呢？这里有以下的有关正则性的结果：如果 $F$ 关于其各个变元是 $C^2$ 的，而且适合以下的勒让德(Legendre)条件 $\det[\mathrm{Hess}_p F(t,u,p)]\neq 0$，则若 $u\in M$ 使 $J(u)\to\min.$，$u$ 必属于 $C^2$，而且在古典意义下(即按通常微积分教本的意义)满足 E—L 方程组(8). 这个结果并不困难. 有的书上把它归功于希尔伯特，而希尔伯特本人与柯朗特合写的名著《数学物理方法，I》第四章中又将它归功于杜•波瓦•雷蒙，且证法完全相同(R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol. 1. p. 200)，但是只限于一个自变量 $t$. 多个自变量时正则性问题是一个十分严重的问题，下面讲到狄利克雷原理时会稍加介绍. 但无论如何，我们不把注意力放在这类问题上，而来看一下

捷线问题是怎样解的.

前面已说过,对于捷线问题,由(1)

$$J=\int_0^{x_B}\left(\frac{1+[f'(x)]^2}{2gf(x)}\right)^{\frac{1}{2}}\mathrm{d}x.$$

于是相应的 $F$ 是

$$F(t,u,\dot{u})=\sqrt{1+[f'(x)]^2}/\sqrt{2gf(x)}. \tag{9}$$

不过 $(t,u,\dot{u})$ 换成了 $(x,f(x),f'(x))$.

一般说来,求解一个 E-L 方程并非易事,不过在(9)这个特例下 $F$ 不显含自变量 $t$(即现在的 $x$).这样,我们可以得出

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\dot{u}F_p-F)&=\ddot{u}F_p+\dot{u}\frac{\mathrm{d}}{\mathrm{d}t}F_p-F_u\dot{u}-F_p\ddot{u}\\&=\dot{u}\left(\frac{\mathrm{d}}{\mathrm{d}t}F_p-F_u\right)=0.\end{aligned}$$

因此

$$F(u,\dot{u})-\dot{u}F_p(u,\dot{u})=C. \tag{10}$$

有一些解常微方程经验的读者都会知道,我们得出了方程(8)的一个初积分或称首次积分.每一个初积分就是一个守恒律.有了一个初积分就可以把方程降低一阶.例如(10)就表明其左方的函数(这是一个物理量)在 E—L 方程的积分曲线(这是一个轨道 $t\to u(t)$)上守恒.因此,找出一个初积分在物理上与数学上有重大的意义.在§1 中我们就指出,开普勒的三定律就是关于太阳和一个行星的牛顿运动方程的解及其性质.其实,求解这组运动方程的方法也是去找它的初积分.其中之一是面积速度(开普勒第二定律),它就是角动量守恒定律(§1.定理 2),然后再利用定理 4 所表述的能量守恒.那么怎样看出有守恒律呢?关键在于 $F$ 在某些变换下有不变性,即对称性.这个例子中 $F$ 中不显含 $t$ 就是一种对称性.发现一个力学系统的守恒律与对称性的联系是德国女数学家爱米•诺特(E.Nöther,1882—1935)的卓越贡献.

再回到(10).把 $F$ 的表达式代入其中,把 $2g$ 并入常数 $C$,并且改一下记号,我们有(用 $y$ 表 $f(x)$)

$$\frac{\sqrt{1+y'^2}}{\sqrt{y}}-\frac{y'^2}{\sqrt{y}\sqrt{1+y'^2}}=\frac{1}{\sqrt{C_1}}$$

化简后有

$$y'^2=\frac{C_1-y}{y}.$$

作一个变换 $y=\frac{C_1}{2}(1-\cos u)$,$y'=\frac{C_1}{2}u'\sin u$,代入上式得

$$\frac{C_1}{2}(1-\cos u)\mathrm{d}u=\pm\mathrm{d}x,$$

对 $u$ 积分后有

$$x=\frac{C_1}{2}(u-\sin u)+C_2,\ y=\frac{C_1}{2}(1-\cos u).$$

我们要求捷线过 $A$（原点）与 $B$. 为使它过原点，应有 $C_2=0$，故

$$x=\frac{C_1}{2}(u-\sin u),\ y=\frac{C_1}{2}(1-\cos u).\tag{11}$$

$C_1$ 由 $B$ 之坐标决定，但是(11)正是以 $u$ 为参数的旋轮线(cycloid)方程.

旋轮线是一个非常重要的曲线. 在大约与伯努利研究捷线问题差不多的时间，惠更斯也研究过所谓等时曲线(tautochrone)问题，所谓等时曲线 $L$ 是这样一条曲线：任一质点在重力作用下沿 $L$ 下滑，不论起点 $A$ 位置何在，滑到最低点 $B$ 所需的时间总是相同的(图 3-6-3). 惠更斯证明了，等时曲线也一定是旋轮线，正是因为这个原因，旋轮线又称摆线. 旋轮线还有许多其它的有趣性质.

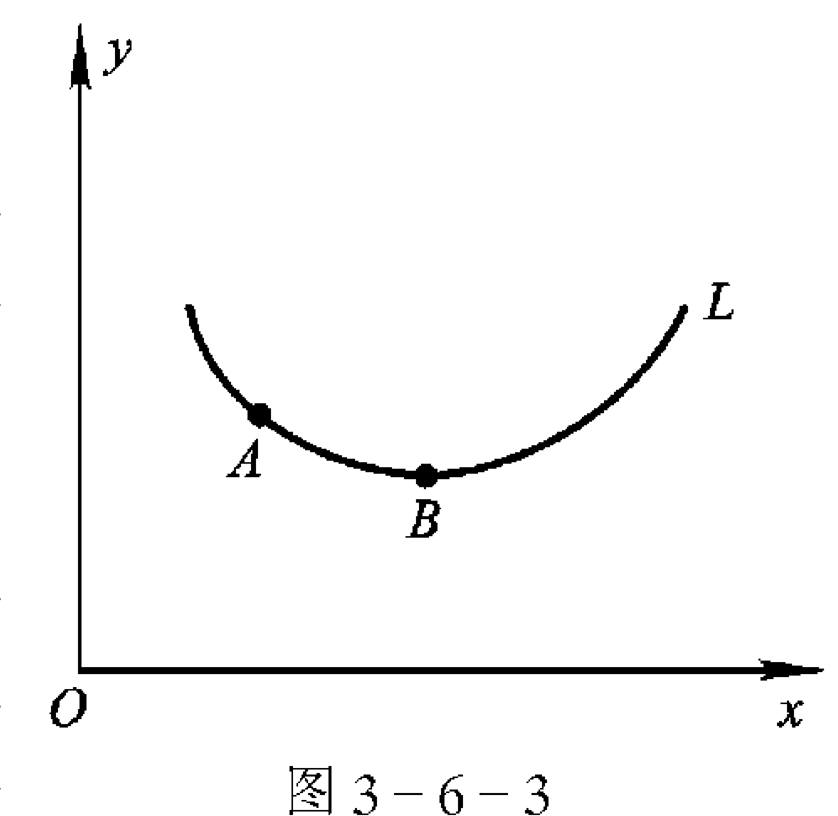


图 3-6-3

**3. 最小作用原理**　变分学中有许多很细致的问题，需要作相当艰难的数学处理. 下面我们将不涉及这类问题，而介绍它在数学其它分支与数学物理中的一些应用. 对每一个应用我们又都只能开一个头，因为深入下去一步就会进入柳暗花明又一村的境地，只有求助于专门著作了.

首先是力学系统的基本规律可以用变分原理来表述. 力学系统在我们这里就指一个质点组，这个质点组可以是有限的，例如有 $N$ 个质点，第 $i$ 个质点的质量为 $m_i$，位置为 $(x_i(t),y_i(t),z_i(t))$，速度为 $(\dot{x}_i(t),\dot{y}_i(t),\dot{z}_i(t))$. 在每一时刻，这 $N$ 个质点的位置可以用 $\mathbf{R}_x^{3N}$ 中的一个点 $(x_1,y_1,z_1;\cdots;x_N,y_N,z_N)$ 来表示. 它们的集合称为构形空间(configuration space). 它们的速度也是另一个 $\mathbf{R}_v^{3N}$ 空间中的一个点 $(u_1,v_1,w_1;\cdots;u_N,v_N,w_N)$. 我们称 $\mathbf{R}_x^{3N}\times\mathbf{R}_v^{3N}$ 为这个质点组的相空间(phase space). 注意，时间总被认为是一个独立的参数，而不进入相空间中. 但是这个质点组可能受一定的约束. 力学中的约束是多种多样的. 例如考虑一个骰子在碗里的运动，如果把骰子看成一个质点，则其三个坐标不会是独立的而要适合这个碗的曲面方程. 所以，一般情况下，一个力学系统的构形空间不一定是 $\mathbf{R}_x^{3N}$，而可以是一个维数较低的“曲面”(准确些说是一个微分流形) $M$. 它的各点的局部坐标我们写成 $(q_1,\cdots,q_N)$，称为广义坐标，而 $(\dot{q}_1(t),\cdots,\dot{q}_N(t))$ 称为广义速度. 广义速度之空间(相应于 $\mathbf{R}_v^{3N}$)就是 $M$ 的切空间，这些概念我们将在第七章中详细讨论. 但不论如何，时间 $t$ 总是在相空间以外的独立参数. 需要提醒的只有一件事，在 $\mathbf{R}_x^{3N}$ 的情况下，例如 $(x_1,y_1,z_1)$ 都是同一个质点的三个坐标，因而相应于同一个 $m_1$. 现在就不一定了. 因此，我们就直接写 $m_1,\cdots,m_N$(如果有必要的话)，而 $m_i$ 与 $m_j$ 可能代表同一个质点的质量，也可能代表不同质点的质量.

牛顿把整个力学建筑在他的三个定律上. 而且，正如他在《原理》一书中所说的，他要尽可能完全地仿照欧几里得的方法论，所以专门在全书之首写了两节，一是“定义”. “定义”了何谓物质的量(即质量)、运动(其实是动量)、惯性、外力……然后第二节是“运动的公理或定律”，把三大定律分别记为定律Ⅰ，Ⅱ，Ⅲ，即全部公理只有这三条. 其它一切，甚至力的平行四边形定律都是“推论”. 牛顿说：“迄今为止我叙述的原理已为数学家所接受，也得到大量实验的验证”. 因此，其“真理性”应是无可置疑的(其实牛顿并不那么简单从事，他本人对这个体系中的许多问题，如绝对空

间、引力的超距作用有很深的怀疑)，所以接下来的工作也就如欧几里得一样，用逻辑方法推论出一个又一个定理.从这个意义上说，牛顿的方法论是"先验的(a priori).

莫培督的方法是另外一回事.他的出发点之一是，既然光的传播服从某个变分原理(费马原理)，则粒子的运动也可能有一个相应的变分原理，即最小作用原理.欧洲大陆上的数学与物理学家对于牛顿的体系是激烈反对的，他们追随的是笛卡儿，而牛顿反对笛卡儿也是十分明确的.尽管牛顿承认自己只不过是站在巨人的肩上，而笛卡儿则是巨人之一.这些争论有深刻的哲学含意，而远非优先权之争.当"知识产权"之争是把"知识"放在"产权"之前时，是会给我们带来丰硕成果的.而放错了地位，甚至只有"产权"而并无"知识"，那就是另一回事了.莫培督说自己的最小作用原理是"由一个形而上学的原理得出的关于运动和静止的法则".可是这个原理的真理性绝不是用哲学的思辨所能解决的，而需要看它能否解决具体问题.从这个意义上说，变分原理的应用是"后验"(a posteriori)的.

现在给出这个原理一个明确的说法.以下表述引自朗道和栗弗席兹的名著《力学》(中文译本高等教育出版社 1959 年版，第 2 页)："每一力学体系由一定的函数

$$L(q_1,\cdots,q_N;\dot{q}_1,\cdots,\dot{q}_N;t) \tag{12}$$

(或简写为 $L(q,\dot{q},t)$)来描述，而体系的运动满足下面的条件：

假定在 $t=t_1$ 和 $t=t_2$ 的时刻，体系占有两个确定的位置，这两个位置分别由两组坐标值 $q^{(1)}$ 和 $q^{(2)}$ 决定.这时，体系在两个位置之间按照使积分

$$S=\int_{t_1}^{t_2} L(q(t),\dot{q}(t),t)\mathrm{d}t \tag{13}$$

有最小可能值的方式运动.函数 $L$ 叫做该体系的拉格朗日函数(Lagrangian)，而积分(13)则叫做作用量(action)".

这里要提醒一下，上文说运动轨道使 $S$ 取最小值，实际上则只使 $S$ 取平稳值.若是 $|t_1-t_2|$ 充分小，则变分学的一般理论可以证明，平稳值一定是局部的极小.

由此可见，求解一个力学系统的运动问题，归结为 $S$ 在边值条件 $q|_{t=t_i}=q^{(i)}$，$i=1,2$ 下的极值问题.由定理 2 有

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_j}\right)-\frac{\partial L}{\partial q_j}=0,\ j=1,2,\cdots,N, \tag{14}$$
$$q_j(t_i)=q_j^{(i)},\qquad i=1,2.$$

(14)称为拉格朗日方程.从数学上说，问题已经完全清楚了，但是从物理上说，拉格朗日函数究竟是什么还完全不清楚，甚至什么是质量，什么是能量也完全没有谈.在拉格朗日改进莫培督的工作时，他从一些特例发现，

$$L=T-V, \tag{15}$$

$T$ 与 $V$ 分别是动能与位能.而在朗道的著作中则从时间与空间的均匀性以及运动方程在任何惯性系中应该有相同形式出发，首先指出，对于一个自由粒子

$$L=\alpha\dot{v}^2=\alpha(\dot{x}^2+\dot{y}^2+\dot{z}^2).$$

于是记 $\alpha=\frac{m}{2}$，并称 $m$ 为质量，$\frac{1}{2}m\dot{v}^2$ 为动能.如果是考虑质点组，若一个力学系统由 $A$ 和 $B$

组成，则 $L$ 不但只是 $L_A+L_B$，还要加一项表示二者之间的相互作用，这个相互作用显然只应与 $A$，$B$ 之位置相关，这样就会得到

$$L=\sum_{i=1}^{N}\frac{m_i}{2}(\dot{x}_i^2+\dot{y}_i^2+\dot{z}_i^2)-V(x). \tag{16}$$

而且称 $V(x)$ 为位能. 同时也证实了 $m_i$ 作为第 $i$ 个质点的质量的合理性. 但是不论怎么说，这些论据都不如牛顿的三个“公理”那么有说服力. 所以到最后还是要看是否与我们已知的结果一致. 以(16)代入(14)，并注意现在 $q_j$ 变成了 $(x_i,y_i,z_i)$ 等等，即得拉格朗日方程

$$\begin{aligned} m_i\ddot{x}_i&=-\frac{\partial V}{\partial x_i},\\ m_i\dot{y}_i&=-\frac{\partial V}{\partial y_i},\\ m_i\ddot{z}_i&=-\frac{\partial V}{\partial z_i}. \end{aligned} \tag{17}$$

这恰好是牛顿的运动方程. 到现在，我们前面称之为质量、动能、位能的物理量与牛顿的经典方法给出的是一致的. 不过现在的附加条件不是初始条件，而是边值条件. 这就后验地证明了最小作用原理的重要性.

读者会问，为什么现在出现的不是总能量 $T+V$ 而是动能位能差 $T-V$？实际上，若转到对偶的表述，将得到另一个极重要的物理量——哈密顿函数(Hamiltonian)

$$H=T+V.$$

那么，由能量守恒岂非 $H\equiv\text{const.}$ 吗？这是一个误解. 所谓能量守恒是指这个力学系在某一轨道上总能量不变，而在不同的轨道上，$H$ 之值并不相同. 哈密顿函数并不满足拉格朗日方程，而是满足一组与它对偶的哈密顿方程，这就构成了哈密顿力学. 其基础仍是最小作用原理.

由此可见，变分原理把力学从牛顿的经典的研究方法扩展为拉格朗日力学和哈密顿力学. 这是现代数学伟大力量的一个范例. 我们在这里当然不可能作进一步介绍，但是建议读者去读阿诺尔德的名著《经典力学的数学方法》(V. I. Arnold, Mathematical Methods of Classical Mechanics, 1978，中译本高教出版社，1992 年出版).

可是更重要的是，这个框架适用于许多物理学分支，例如电动力学、量子场论、规范场论等等. 这时这种脱离具体感性经验然后再后验地加以论证的方法更加重要. 但是我们现在不去讨论这些问题了. 用拉格朗日方程代替牛顿的运动方程至少有一个好处，即很容易得到其它坐标系下的运动方程. 例如设有一个质点，质量为 $m$，在位能为 $V(x)$ 的力场中运动. 这时用极坐标去处理，有 $q=(r,\theta)$，而笛卡儿坐标 $x=r\cos\theta$，$y=r\sin\theta$. 现在

$$\begin{aligned} T&=\frac{m}{2}(\dot{x}^2+\dot{y}^2)=\frac{m}{2}[(\dot{r}\cos\theta-r\dot{\theta}\sin\theta)^2+(\dot{r}\sin\theta+r\dot{\theta}\cos\theta)^2]\\ &=\frac{m}{2}(\dot{r}^2+r^2\dot{\theta}^2). \end{aligned}$$

因此，由(15)式，$L=T-V$，所以 E—L 方程为

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{r}}+\frac{\partial L}{\partial r}\equiv m\ddot{r}-mr\dot{\theta}^2+\frac{\partial V}{\partial r}=0,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\theta}}+\frac{\partial L}{\partial \theta}\equiv m\frac{\mathrm{d}}{\mathrm{d}t}(r^2\dot{\theta})+\frac{\partial V}{\partial \theta}=0.$$

特别重要的是位能 $V$ 不依赖于 $\theta$，这时我们得到有心力场，而上面第二个方程成为 $\frac{\mathrm{d}}{\mathrm{d}t}(r^2\dot{\theta})=0$ 亦即

$$r^2\dot{\theta}=\mathrm{const}.$$

但是这就是面积速度(亦即角动量)守恒，所以我们十分简洁地得到了开普勒第二定律.

我们要介绍一个重要的概念. 即定义一个力学系统的广义动量是 $p=\left(\frac{\partial L}{\partial \dot{q}_1},\cdots,\frac{\partial L}{\partial \dot{q}_N}\right)=\mathbf{grad}_{\dot{q}}L$. 这与牛顿力学中动量 = 质量 × 速度不同，广义动量 ≠ 质量 × 广义速度. 如果拉格朗日函数 $L=T-V$，而动能 $T$ 是 $\dot{q}$ 的二次型：$T=\sum_{i,j=1}^{N}g_{ij}(q)\dot{q}_i\dot{q}_j$，$g_{ij}=g_{ji}$，$V=V(q)$ 与 $\dot{q}$ 无关(我们只看 $L$ 不显含 $t$ 的情况)，则广义动量 $p=(p_1,\cdots,p_N)$ 现在成为

$$p_i=\sum_j g_{ij}\dot{q}_j. \tag{18}$$

考虑到质量已含于 $g_{ij}$ 中，则若记 $(g^{ij})=(g_{ij})^{-1}$，我们会有

$$\dot{q}_j=\sum_i g^{ij}p_i. \tag{19}$$

由此可见，仅当 $(g_{ij})$ 是对角矩阵，即 $(g_{ij})=\begin{pmatrix}m_1 & & \\ & \ddots & \\ & & m_N\end{pmatrix}$ 时，我们才能得到与

$$\text{动量}=\text{质量}\times\text{速度}，\text{速度}=\frac{1}{\text{质量}}\times\text{动量}$$

“一样”的关系式

$$p_i=m_i\dot{q}_i，\dot{q}_i=\frac{1}{m_i}p_i， \tag{20}$$

否则只能得到矩阵关系式(18)、(19). 实际上，广义动量 $p$ 与广义速度 $\dot{q}$、广义坐标 $q$ 是性质不同的向量，二者之间有一个对偶关系. 我们将在第七章中详细讨论这个区别. 这样，要描述一个力学系统应该给出 $\dot{q}$ 与 $\dot{p}$ 用 $(p,q)$ 的表达式. 这正是哈密顿力学需要解决的问题. 现在我们只看到一个事实：上面的例子中 $L$ 中不含 $\theta$，于是相应于 $\theta$ 的广义动量分量 $\frac{\partial L}{\partial \dot{\theta}}=mr^2\dot{\theta}$ 守恒，从而得到角动量守恒. 一般说来，若拉格朗日函数 $L$ 中不含某一个广义坐标例如 $q_j$，则拉格朗日方程将给出 $\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j}=0$，从而相应于 $q_j$ 的广义动量必守恒. 这种 $L$ 中不显含的坐标称为循环坐标(cyclic coordinates)，只要找到了一个循环坐标就会得到一个守恒律 $p_j=\mathrm{const}$. 前面已经说过，找到一个力学系统的守恒律对研究这个力学系统有重大的意义. 这样一些在牛顿力学中不易得到的关系，在以最小作用原理为基础的拉格朗日力学与哈密顿力学中却是显而易见的结论. 尤其重要的是：这

些结果在物理学其它分支中同样很重要.

**4. 测地线** 现在提出另一个问题:在曲面 $S$ 上有两点 $A,B$ 怎样找出连接 $A,B$ 两点的最短的曲线?首先的问题是,曲面 $S$ 上如何定义长度?仿照上一节,我们把曲面规定为由源空间 $\mathbf{R}^m$ 中的开集 $\Omega$ 到靶空间 $\mathbf{R}^n$ 中的一个光滑的映射,如果用局部坐标来表示,就是

$$\varphi: y_\alpha = \varphi_\alpha(x_1, \cdots, x_m), \alpha = 1, 2, \cdots, n. \tag{21}$$

以下我们规定:若用拉丁字母如 $i,j$ 等作下标,总是由 1 变到 $m$ 而与源空间中的坐标维数一致;若用希腊字母如 $\alpha,\beta$ 等作下标,则由 1 变到 $n$ 而与靶空间的坐标维数一致. (21)适合最大秩条件. 我们还规定这些函数属于 $C^2(\Omega)$,而不去设法减少其光滑性的要求. 在靶空间中,弧长

$$\mathrm{d}s^2 = \sum_{\alpha=1}^n \mathrm{d}y_\alpha^2,$$

如果限制在 $S = \varphi(\Omega)$ 上,就给出了其上的弧长为

$$\begin{aligned}\mathrm{d}s^2 &= \sum_{\alpha=1}^n \sum_{i,j=1}^m \frac{\partial \varphi_\alpha}{\partial x_i}\mathrm{d}x_i \cdot \frac{\partial \varphi_\alpha}{\partial x_j}\mathrm{d}x_j = \sum_{\alpha=1}^n \sum_{i,j=1}^m \left(\frac{\partial \varphi_\alpha}{\partial x_i}\frac{\partial \varphi_\alpha}{\partial x_j}\right)\mathrm{d}x_i \mathrm{d}x_j \\ &= \sum_{i,j=1}^m g_{ij}\mathrm{d}x_i \mathrm{d}x_j,\end{aligned} \tag{22}$$

这里

$$g_{ij}(x) = \sum_{\alpha=1}^n \frac{\partial \varphi_\alpha}{\partial x_i}\frac{\partial \varphi_\alpha}{\partial x_j}. \tag{23}$$

于是 $g_{ij}(x) = g_{ji}(x)$,而且矩阵 $(g_{ij}(x))$ 对任一点 $x \in \Omega$ 都是正定的. 这件事并不需要"硬算". 因为 $\mathrm{d}s^2 \geqslant 0$,而且 $\mathrm{d}s^2 = 0$ 当且仅当一切 $\mathrm{d}y_\alpha = 0$,但由最大秩假设 $\mathrm{d}y_\alpha = 0$ 当且仅当一切 $\mathrm{d}x_i = 0$. 所以(22)当且仅当一切 $\mathrm{d}x_i = 0$ 时为 0. 即是说,作为 $\mathrm{d}x_i$ 的二次型 $\mathrm{d}s^2$ 是恒正的,亦即 $(g_{ij}(x))$ 是正定矩阵. 这样它就有逆矩阵存在. 我们规定把逆矩阵记为 $(g_{ij}(x))^{-1} = (g^{ij}(x))$.

这样我们看到,在 $S$ 上现在有了一个度量(22). 它是由靶空间(即 $S$ 的包含空间)中的欧几里得度量 $\mathrm{d}s^2 = \sum_{\alpha=1}^n \mathrm{d}y_\alpha^2$ 诱导而来的. 但是从现在起我们可以完全"忘记"诱导它的根源(映射 $\varphi$),而只需要记住两点:1. 它是对称的:$g_{ij} = g_{ji}$;2. $(g_{ij}(x))$ 是正定的. 换句话说,设 $S$ 是一个"曲面"(微分流形),其上用(22)式定义一个"度量",而且 $(g_{ij})$ 适合以上的要求. 因为 $S$ 并不一定嵌入或浸入在某一个包含空间 $\mathbf{R}^n$ 内,所以也就没有一个映射(21)使得(23)式成立. 具有适合这两个条件的度量的微分流形称为一个黎曼流形,我们要在 $S$ 上确定如上所述的连接 $A,B$ 两点的弧. 关于微分流形的准确意义详见第七章,我们现在就只讨论如何用变分学来解决以上问题.

现在再换一个角度来讨论它. 从几何上看,连接 $A,B$ 两点的曲线 $C: x = x(t)$ 的弧长是一个泛函

$$J_{AR}(C) = \int_A^B \mathrm{d}s = \int_a^b \sqrt{\sum_{i,j=1}^m g_{ij}(x)\frac{\mathrm{d}x_i}{\mathrm{d}\sigma} \cdot \frac{\mathrm{d}x_j}{\mathrm{d}\sigma}}\,\mathrm{d}\sigma. \tag{24}$$

这里 $\sigma$ 是曲线 $C$ 的参数. 如果选 $\sigma$ 为时间 $t$,并设有一个单位质量的自由质点沿 $C$ 运动,则

$$\frac{1}{2}\sum_{i,j=1}^m g_{ij}(x)\dot{x}_i \dot{x}_j \mathrm{d}t = \frac{1}{2} \cdot 1 \cdot \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = \frac{1}{2}mv^2$$

是动能,而

$$J_{ER}(C)=\frac{1}{2}\int_a^b\sum_{i,j=1}^m g_{ij}(x)\dot{x}_i\dot{x}_j\mathrm{d}t,\dot{x}_i=\frac{\mathrm{d}x_i}{\mathrm{d}t} \tag{25}$$

就是动能对时间的积分$\int_a^b T\mathrm{d}t$，也就是莫培督的作用量.上一个例子中说过，所谓位能是一个质点组中各质点相互位置产生的能量，现在既然是讨论一个自由质点的运动，自然就没有位能 $V$ 了，或者说 $V=0$.所以现在拉格朗日函数 $L=T-V=T$，而(25)就是拉格朗日的作用量.现在要问，(24)的极值曲线即使 $J_{AR}(C)\to\min$ 的曲线 $C$ 与(25)的极值曲线有何关系？若记沿 $C$ 的弧长与能量分别为 $L(C)$ 与 $E(C)$，我们有

**引理 3** 若 $C$ 是一条 $C^1$ 曲线.则

$$L(C)\leqslant\sqrt{2(b-a)}\sqrt{E(C)}. \tag{26}$$

这里的等号当且仅当$\sqrt{\sum g_{ij}(x)\dot{x}_i\dot{x}_j}\equiv 1$ 时成立.

证 由施瓦茨不等式

$$L(C)=\int_a^b 1\cdot\sqrt{\sum g_{ij}(x)\dot{x}_i\dot{x}_j}\,\mathrm{d}t\leqslant\left(2\int_a^b 1\mathrm{d}t\right)^{\frac{1}{2}}\left(\frac{1}{2}\int_a^b\sum g_{ij}(x)\dot{x}_i\dot{x}_j\mathrm{d}t\right)^{\frac{1}{2}}$$

$$=\sqrt{2(b-a)}\cdot\sqrt{E(C)}.$$

于是等号成立时，参数 $t$ 恰好就是弧长.这时 $L(C)=b-a$，于是(26)给出

$$L(C)=\sqrt{2L(C)}\cdot\sqrt{E(C)}\text{即 }L(C)=2E(C). \tag{27}$$

引理 3 中我们限制了只讨论 $C^1$ 曲线.前面在讲到可容许函数集时，总是作这样的假设的.

引理 3 告诉我们，为了求 $L(C)$ 的极值曲线，最好是求 $E(C)$ 的极值曲线，因为在以弧长为参数时，能量的极值即弧长之极值.这样，我们就可以对 $E(C)$ 求其 E－L 方程如下：

$$\frac{\partial}{\partial\dot{x}_k}\left(\sum_{i,j=1}^m g_{ij}\dot{x}_i\dot{x}_j\right)=2\sum_{i=1}^m g_{ik}\dot{x}_i\qquad(\text{注意 } g_{ik}=g_{ki}),$$

$$\frac{\partial}{\partial x_k}\left(\sum_{i,j=1}^m g_{ij}\dot{x}_i\dot{x}_j\right)=\sum_{i,j=1}^m\frac{\partial g_{ij}}{\partial x_k}\dot{x}_i\dot{x}_j.$$

注意到 $g_{ij}(x)$中不含 $t$，就有

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial}{\partial\dot{x}_k}\sum_{i,j=1}^m g_{ij}\dot{x}_i\dot{x}_j\right]=2\sum_{i,j=1}^m\frac{\partial g_{ik}}{\partial x_j}\dot{x}_i\dot{x}_j+2\sum_{i=1}^m g_{ik}\ddot{x}_i.$$

和前面一样，用$(g^{ij})$表示$(g_{ij})^{-1}$，则 E－L 方程成为

$$0=\ddot{x}_l+\frac{1}{2}\sum_{i,j,k=1}^m g^{kl}\left(2\frac{\partial g_{ik}}{\partial x_j}-\frac{\partial g_{ij}}{\partial x_k}\right)\dot{x}_i\dot{x}_j$$

$$=\ddot{x}_l+\frac{1}{2}\sum_{i,j,k=1}^m g^{kl}\left(\frac{\partial g_{ik}}{\partial x_j}+\frac{\partial g_{jk}}{\partial x_i}-\frac{\partial g_{ij}}{\partial x_k}\right)\dot{x}_i\dot{x}_j.$$

引入著名的克里斯托费尔(Christoffel)记号

$$\Gamma_{ij}^l=\frac{1}{2}\sum_{k=1}^m g^{kl}\left(\frac{\partial g_{ik}}{\partial x_j}+\frac{\partial g_{jk}}{\partial x_i}-\frac{\partial g_{ij}}{\partial x_k}\right), \tag{28}$$

则 E－L 方程成为

$$\ddot{x}_l + \sum_{i,j=1}^{m} \Gamma_{ij}^l \dot{x}_i \dot{x}_j = 0, l = 1, 2, \cdots, m. \tag{29}$$

特别是，若取参数 $t$ 为弧长 $s$，上式成为

$$\frac{\mathrm{d}^2 x_l}{\mathrm{d}s^2} + \sum_{i,j=1}^{m} \Gamma_{ij}^l \frac{\mathrm{d}x_i}{\mathrm{d}s} \frac{\mathrm{d}x_j}{\mathrm{d}s} = 0, l = 1, 2, \cdots, m, \tag{30}$$

它还是 $L(C)$ 的 E－L 方程. 我们给出

**定义 1** (30)的解称为 $S$ 关于度量 $(g_{ij})$ 的测地线. (29)则称为关于能量的测地线.

当 $t$ 即 $s$ 时，这两种测地线是一致的.

**5. 狄利克雷原理** (Dirichlet principle) §4 中讲到解析函数时曾提到黎曼的解析函数理论是一个几何理论，而它是与黎曼关于静电场理论的研究密切相关的. 如果我们把牛顿关于力学的理论看成人类对自然规律的认识，特别是物理学的第一次伟大的综合，则麦克斯韦的电磁理论毫无愧色的是物理学的第二次大综合. 也正如牛顿力学是一个时代的产物一样，在麦克斯韦之前，许多数学家和物理学家都极为关注电磁现象，其中就包括黎曼. 黎曼甚至考虑过电与引力的统一问题. 当时，黎曼关心的是静电场问题，这与法拉第—麦克斯韦的电磁理论(其中涉及电场和磁场的动力学问题)是不同的. 关于静电场，黎曼时常引述狄利克雷的如下论述：平面区域 $\Omega$ 中的拉普拉斯方程 $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ 的适合边值条件 $u\big|_{\partial\Omega} = g$ 的解，必使 $\iint_{\Omega} \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right]\mathrm{d}x\mathrm{d}y \to \min$ (准确些说是平稳值). 这个积分(现在称为狄利克雷积分)表示静电场的位能，而从物理的观点看来，如果在 $\partial\Omega$ 上放置一些电荷而使电场的电位 $u$ 适合 $u|_{\partial\Omega} = g$，则这些电荷的分布必然是使位能达到极小时才能稳定下来，而一旦电场稳定了以后，它必然适合 $\Delta u = 0$. 其实在 19 世纪中叶就已经有许多数学家和物理学家理解到这一点了，狄利克雷并非第一人. 狄利克雷原理的称号是黎曼给的.

现在我们来证明狄利克雷积分

$$D(u) = \iint_{\Omega} \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right]\mathrm{d}x\mathrm{d}y \tag{31}$$

在可容许函数类

$$M = \{u, u \text{ 在 } \bar{\Omega} \text{ 内适当光滑}, u|_{\partial\Omega} = g\}$$

中的 E－L 方程正是拉普拉斯方程. 和前面一样，考虑

$$D(u + \lambda\varphi) = \iint_{\Omega} [(u_x + \lambda\varphi_x)^2 + (u_y + \lambda\varphi_y)^2]\mathrm{d}x\mathrm{d}y,$$

这里 $\varphi|_{\partial\Omega} = 0$，从而 $u + \lambda\varphi \in M$. 把此式按 $\lambda$ 展开，并且只考虑 $\lambda$ 的线性项，而且令它为 0(也和前面一样，这就是略去高阶无穷小)，有

$$\iint_{\Omega} (u_x\varphi_x + u_y\varphi_y)\mathrm{d}x\mathrm{d}y = 0.$$

应用格林公式，有

$$\iint_{\Omega} (u_x\varphi_x + u_y\varphi_y)\mathrm{d}x\mathrm{d}y = \int_{\partial\Omega} [u_x\cos(n, x) + u_y\cos(n, y)]\varphi\mathrm{d}x - \iint_{\Omega} \Delta u \cdot \varphi\mathrm{d}x\mathrm{d}y.$$

但是 $\varphi|_{\partial\Omega} = 0$，而 $\partial\Omega$ 上的积分为 0，所以对一切适合此条件的 $\varphi$ 均有

$$\iint_{\Omega} \Delta u \cdot \varphi \mathrm{d}x\mathrm{d}y = 0.$$

用变分法的基本引理(它对高维情况也是成立的)即得

$$\Delta u = 0. \tag{32}$$

所以上述变分问题的解也就给出了(32)的狄利克雷问题(就是再附加边值条件 $u|_{\partial\Omega}=g$)的解.

以上我们对 $M$ 中的函数所应有的光滑性以及 $\partial\Omega$ 的光滑性都有意地模糊了,而其实这正是关键所在.

那么如何去找 $u_0\in M$ 使 $D(u_0)\to\min$ 呢? 从(31)看到 $D(u)\geqslant 0$,所以 $\inf\limits_M D(u)$ 必存在.可以找到一个 $u_0\in M$ 使 $D(u_0)=\inf\limits_M D(u)$,则 $u_0$ 即合于所求.这个证明在物理上极为令人信服,而且它使我们不必去求偏微分方程(32)的狄利克雷问题的解,而采用以下的所谓"直接方法"(direct method).在 $M$ 中找这样一个序列 $\{u_k\}$,使 $D(u_k)\to\inf\limits_M D(u)$.这样的 $\{u_k\}$ 称为极小化序列.然后考虑这个序列的极限.这样一个程序不但有理论上的意义而且在实际计算上十分有用.它有许多变化,以保证计算简单、收敛迅速等等.许多实用的数值方法例如伽辽金方法(Galerkin method),特别是有限元法(finite element method)都属于变分法这个大范畴.

但是狄利克雷原理很快就遭到了魏尔斯特拉斯的批评.上面说可以找到 $u_0\in M$,使 $D(u_0)=\inf\limits_M D(u)$,即 $D(u)$ 在 $M$ 中的下确界是可以达到的,因此下确界就成了最小值.魏尔斯特拉斯就此给出了反例,大体如下:令

$$I(u)=\int_{-1}^{1} t^2\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)^2\mathrm{d}t,$$

$M=\{u=u(t),\ u(t)$ 在 $[-1,1]$ 上连续可微而且 $u(-1)=-1,\ u(1)=1\}$.当然 $I(u)\geqslant 0$.今证 $\inf\limits_M I(u)=0$.任取 $\varepsilon>0$ 并令 $u(t)=u_\varepsilon(t)=\arctan\dfrac{t}{\varepsilon}\Big/\arctan\dfrac{1}{\varepsilon}$.很明显 $u_\varepsilon(t)\in M$,$\dfrac{\mathrm{d}u_\varepsilon}{\mathrm{d}t}=\dfrac{1}{\varepsilon\arctan\dfrac{1}{\varepsilon}}\cdot\dfrac{1}{1+\left(\dfrac{t}{\varepsilon}\right)^2}=\dfrac{\varepsilon}{\arctan\dfrac{1}{\varepsilon}}\dfrac{1}{t^2+\varepsilon^2}$,所以当 $\varepsilon\to 0$ 时,

$$\begin{aligned}I(u_\varepsilon)&<\int_{-1}^{1}(t^2+\varepsilon^2)\left(\frac{\mathrm{d}u_\varepsilon}{\mathrm{d}t}\right)^2=\frac{\varepsilon^2}{\left(\arctan\dfrac{1}{\varepsilon}\right)^2}\int_{-1}^{1}\frac{\mathrm{d}t}{t^2+\varepsilon^2}\\&=2\varepsilon/\arctan\frac{1}{\varepsilon}\to 0.\end{aligned}$$

这就表明 $\inf\limits_M I(u)=0$.但这个下确界是不能达到的.因为若 $I(u_0)=0$,应有 $\int_{-1}^{1} t^2\left(\dfrac{\mathrm{d}u_0}{\mathrm{d}t}\right)^2=0$.因此 $\left(\dfrac{\mathrm{d}u_0}{\mathrm{d}t}\right)^2\equiv 0$ 而 $u_0\equiv\mathrm{const}$.但是常值函数不可能适合边值条件 $u(\pm 1)=\pm 1$.

所谓直接方法还有另一个困难,下面是阿达玛举出的例子:设 $\Omega$ 是单位圆 $x^2+y^2<1$.$M$ 中的边值条件是

$$u|_{\partial\Omega}=u(1,\theta)=\sum_{n=1}^{\infty}\frac{1}{n^2}\cos(n!\ \theta). \tag{33}$$

上式右方的级数是一致收敛的,因此代表一个连续函数.

这个狄利克雷问题很容易求解.因为(33)右方是

$$\sum_{n=1}^{\infty}\frac{1}{n^2}[\cos(n!\ \theta)+i\sin(n!\ \theta)]=\sum_{n=1}^{\infty}\frac{1}{n^2}\mathrm{e}^{\mathrm{i}n!\theta}$$

之实部,而这个级数又是

$$\sum_{n=1}^{\infty}\frac{1}{n^2}z^{n!}=\sum_{n=1}^{\infty}\frac{1}{n^2}r^{n!}\mathrm{e}^{\mathrm{i}n!\theta}$$

在 $r=1$,即$\partial\Omega$ 上之边值.这是一个幂级数,它有"许多"系数为 0.(这类级数称为缺项级数(lacunary series),它们所代表的解析函数有许多有趣的性质),它的收敛圆是$|z|<1$.

记其和为 $f(z)$.其实部

$$u(r,\theta)=\sum_{n=1}^{\infty}\frac{1}{n^2}r^{n!}\cos(n!\ \theta)$$

就在 $\Omega$ 内调和,在 $\bar{\Omega}$上连续且适合边值条件(33),即为阿达玛例子之解.现在我们来计算它的狄利克雷积分.注意到,由 C—R 方程 $u_x^2+u_y^2=u_x^2+v_x^2=|u_x+\mathrm{i}v_x|^2=f'(z)\bar{f}'(z)$.所以

$$\begin{aligned}D(u)&=\iint_{|z|<1}|f'(z)|^2\mathrm{d}z=\lim_{\rho\to1}\iint_{|z|\leqslant\rho\leqslant1}f'(z)\cdot\bar{f}'(z)\mathrm{d}z\\&=\lim_{\rho\to1}2\pi\iint_{|z|<\rho}r\sum_{m=1}^{\infty}\frac{(m-1)!}{m}r^{m!-1}\mathrm{e}^{\mathrm{i}(m!-1)\theta}\cdot\sum_{n=1}^{\infty}\frac{(n-1)!}{n}r^{n!-1}\mathrm{e}^{-\mathrm{i}(n!-1)\theta}\mathrm{d}r\mathrm{d}\theta\\&=\lim_{\rho\to1}2\pi\int_0^{\rho}\sum\frac{[(m-1)!]^2}{m^2}r^{2m!-1}\mathrm{d}r=\lim_{\rho\to1}2\pi\sum_{m=1}^{\infty}\frac{[(m-1)!]^2}{2m!\ m^2}\rho^{2m!}\\&=2\pi\sum_{m=1}^{\infty}\frac{m!}{2m^4}=+\infty.\end{aligned}$$

这里在对 $\theta$ 积分时要利用$\{\mathrm{e}^{\mathrm{i}k\theta}\}$之正交性而发现只余下 $m=n$ 的各项.要把$\lim\limits_{\rho\to1}$移到求和号下要一些特别的技巧,我们略去了而只写出结果.

这就是告诉我们,这个解没有有限的 $D(u)$.所以不可能从 $\inf D(u)$中找到它.

尽管这个方法有许多缺点,但是它是如此有说服力,所以人们一直努力"挽救"它.到 1901 年才由希尔伯特完全地证明了狄利克雷原理.为此,他附加了一个条件:$M$ 中一定能找到一个函数 $u$ 适合其边值条件,而且使 $D(u)<+\infty$.这显然是针对阿达玛的例子而来的.这里的过程大体如下:先看极小化序列$\{u_k\}$.至少要找到一个收敛的子序列,不妨假设就是$\{u_k\}$,不但使$\lim\limits_{h\to\infty}u_k=u_0$,而且使得$\lim\limits_{k\to\infty}\frac{\partial u_k}{\partial x}=\frac{\partial u_0}{\partial x}$,$\lim\limits_{k\to\infty}\frac{\partial u_k}{\partial y}=\frac{\partial u_0}{\partial y}$,还使得$\lim\limits_{k\to\infty}D(u_k)=D(u_0)$.这就要求在积分号下和微分号下求极限,我们的经验告诉我们这已经是很困难的事.但这还没有完,$D(u_0)$中只出现 $u_0$ 的一阶导数,狄利克雷问题则涉及 $u_0$ 的二阶导数,那么极限函数 $u_0$ 是否有这样高的正则性呢?在最简单的一维变分问题中,我们指出了有希尔伯特的正则性定理.在高维情况下也有这样的结果.但是为了要完满地回答这些问题,我们就要推广导数的理论,推广积分的理论,推广收敛性的理论.可以说是风云际会,在 20 世纪二三十年代一大批数学家,其中突出的有前苏联的索波列夫(S.L.Sobolev),完成了这个任务.可以说这是现代分析数学突出的一章.附带说一下,索波列夫又把自己最重要的结果称为"引理",不过一般人却称之为嵌入定理.这里涉及的问题,我们将在以下几章中部分地讨论.正因为如此,我们现在就对 $M$ 中函数的光滑性与$\partial\Omega$ 的光滑性略去不

讲，而有意地模糊了.

以上我们举了三个例子，看见了变分学所开辟的三个广阔的新天地.可惜，每一个例子的参考文献都是篇幅巨大的大著.

## §7 不可求导的函数

这一节只是一个简单的介绍和一些可以争论的议论.

我们首先讨论两件似乎无关的事：第一件是回顾一下19世纪中分析数学发展的困难是怎样解决的.第二件是19世纪末，20世纪初物理学中的一件大事.

我们时常说19世纪是分析数学中注入了严密性的时代.这个时代的一位有重大贡献的代表性人物阿贝尔在1826年说过，“人们在分析中确实发现了惊人的含糊不清之处.这样一个完全没有计划和体系的分析，竟有那么多人能研究过它，真是奇怪.最坏的是，从来没有严格地对待过分析.非常奇怪的是这种方法只导致极少几个所谓悖论”(引文转引自克莱因《古今数学思想》中译本，第四卷，40章.上海科技出版社，1981年出版).阿贝尔这样说是很自然的，即以函数概念为例，我们认识函数确实是由我们的感觉经验而非严格的公理化定义开始的.但是，这种感觉经验与大自然的规律符合到何种程度？它归根结蒂是不是也只是某种观测仪器给我们的经验曲线？人们总是不自主地想对“函数”加上种种限制.人们最容易接受的首先是连续性，连续性局部地看起来就是不间断和连通.由此自然想到，连续曲线下面的曲边梯形应该有面积$\int_a^b f(x)\mathrm{d}x$，所以人们以为只有可积的函数才是自然的，才是合理的，否则都是病态的.经验告诉我们曲线在每一点都有方向，例如质点运动的轨道总有方向.人们确实“感觉”到了$\frac{\mathrm{d}f}{\mathrm{d}x}$.同样人们也可以“感觉”到$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}$，因为一块石头或者一片树叶掉到头上造成的“感觉”确实是不一样的.于是人们又认为只有光滑函数才不是病态的.所以雅可比把这些光滑的函数称为“合理的函数”.但是人究竟能不能“感觉”到一个函数是$C^4$而不是$C^5$函数？人能不能依据“感觉”来区别$C^\infty$函数与解析函数？实际上，人是越来越多地依靠逻辑，看由此形成的概念会得出什么样合逻辑的结论，看是否能更深刻地揭示自然的规律.“不识庐山真面目，只缘身在此山中”，所以到19世纪的阿贝尔和今天的我们才会在慨叹含糊不清之处多得令人吃惊的同时，也感到奇怪，这么多的含糊不清之处，竟然只形成“极少几个悖论”，竟然没有形成完全错误需要抛弃的理论.但是这种情况终于到了不能不解决的程度.我们下面连续三章都可以说是为了说明这一点.有些定理的叙述和证明将比前几章更细致，以避免出现漏洞.然而，为什么数学科学中“错误”一直较少呢？这不能不归功于在数学中有逻辑推理这样的工具，有严格性这样一种标准.回想一下，二千多年前的欧多克萨斯用不等式绕过了芝诺关于运动的悖论.现在人们发现，根本没有什么必要讨论无穷小，不必用玄学的语言来添乱，只要有极限理论一切都解决了.于是柯西说：$\lim\limits_{x\to a} f(x)=A$就是“当$x$越来越接近$a$时，$f(x)$越来越接近$A$”，而魏尔斯特拉斯还认为“越来越接近”的说法不清楚，仍然留下了一些来自运动学的概念的痕迹，而应该用$\varepsilon-\delta$的语言，把问题归结为不等式.从这一点说，魏尔斯特拉斯

是回到了欧多克萨斯，而把来自运动学的概念和语言诸如时间、运动等等，都“请”出了数学.从这一点讲，“变”量的数学变成了静态的数学.其实这是很自然的，因为例如时间更多的是物理学的研究对象.把时间引入数学就不可避免地带来了与此相关的物理学中的困难.所以到了19世纪末，人们就在这样一种十分讲究严格性的气氛中来讨论数学.其中自然也包括微积分.从表面上看，人们总在作“反面文章”.你说什么东西可以求导，数学家就找一个“反例”说这不一定行.人们想求极限，他又找一个“反例”，说求了极限就会出问题.所以许多定理甚至其陈述也变得几乎看不懂.微积分变得更不近人情了，$\varepsilon-\delta$ 如噩梦似的压在学生们的心头.许多人作了种种努力使数学更“人性化”一些，但是没有一个人主张读牛顿、欧拉的“原汁原味”的微积分，而仍然要读被魏尔斯特拉斯等人“消了毒”的微积分.

终于出现了康托尔(G.Cantor，1845—1918)的集合论.本书有一个界限，即不来讨论集合论的问题.我们只想指出，康托尔一生生活在争论之中.由集合论产生的“悖论”可不能说只有“极少几个”了，而且其影响深远.对于集合论，20世纪的两位数学巨人有截然不同的看法.一位是希尔伯特.他说康托尔的集合论为我们建立了数学的天国乐园，现在谁也不能把我们从这个乐园中赶走.另一位同样伟大的数学家是庞加莱.他在《科学与方法》(1913)一书第二编“数学推理”中说了好长一段话：“逻辑有时造成奇异的东西.半个世纪以来，我们看到了出现一大堆异乎寻常的函数，它们似乎力图尽可能不类似于具有某些效用的普通函数.它们不再有连续性，或者也许有连续性，但却没有导数等等.不仅如此，从逻辑的观点来看，正是这些奇异函数才是最普遍的.……从前，当一个新函数被发明时，正是为了实用的目的，今天为了指出我们祖先推理的错误，才特意发明新函数，人们从来也不能从中得到比这更多的东西.”说白了，这就是说现在人们研究数学为的只是证实他们的爷爷无能而且不严格.总之，数学是否过度强调严格化，公理化……妄想“清洗”直观的直觉，以至于远离对大自然的深刻的探索呢？这样，我们来看第二件事.

如果今天还有人不承认分子和原子的存在，那就是很可笑的事.但是当1897年元旦奥地利物理学家和哲学家马赫(E.Mach，1838—1916)(此人在中国名声不佳)在维也纳科学院一次会议上宣布“我不相信原子存在”时，得到的却不是嘲笑，而是有人反对，有人深思.这是物理学一次大辩论中的一幕.这场大争论终于导致物理学的大发展.问题在于，尽管当时从原子“假说”中已得出了许多正确的，与实验相符的结论，原子终究还是不可捉摸的东西.对于原子是什么，有什么性质，还一直没有“直接”的证据.马赫从他的实证哲学的观点看来，原子的存在还只是一种形而上学的假说，所以应该抛弃.“直接”的证据出现在布朗运动的研究.布朗(Robert Brown)是英国植物学家，他注意到花粉在水面上的无规则运动(1827).对此，只能用分子的碰撞来解释.爱因斯坦(1905)做了许多工作，并且计算出诸如阿伏伽德罗常数和分子的平均自由程，后来又经著名的法国实验物理学家佩韩(Jean Baptiste Perrin)系统的实验证实，原子是否存在的争论，才算尘埃落定.正是这位佩韩在他的名著《原子》一书中说：“利用显微镜观察布朗运动，看到浮在溶液中的每个小颗粒来回运动时，我们被实验的真实完全吸引住了.为了对它的轨迹作切线，需要作连接轨迹上非常接近的两点(即该小颗粒在两邻近时刻的位置)的直线，并找出它的极限位置.可是在我们研究过程中，两次观察的时间间隔无论多小，这直线的方向总是持续地变化着.由这项研究，一个不带成见的观察者只能得到函数没有导数的感觉，而终究得不到有切线的曲线的感觉.”这段话转引自克莱因的名著《高观点下的初等数学》第三卷，湖北教育出版社，1993年，中译本46页.译者是已故的著名数学家吴大任教授和夫人陈□教授.同页上还引述了博雷尔(E.Borel——测

度理论的首创者之一)的一篇评论:"分子理论及其数学",可惜无法找到原文.布朗运动的重要性早就不限于研究花粉的随机活动,而在随机现象的数学与物理理论乃至金融理论中都极为重要,而几乎所有布朗运动的轨道又都是处处不可求导的函数.当然这不是说不可求导函数的研究都正是这些"实际需要"的"期货".这样说,不仅不真实,而是不严肃.到了20世纪70年代末、80年代初,数学中出现了所谓分形几何学的研究,指出处处没有切线的曲线等等,在大自然中是十分普遍而且重要的.分形几何学在数学中也有长久的源流,与许多数学分支有关,所以我们没有力量去涉及有关的问题,而只想指出,正是由庞加莱开创的天体力学的研究中,在研究天体运动轨道的渐近状况时,就会出现构造极为奇特的集合.分形理论与它有密切关系.可见,数学走到了这一步,这种严格化、公理化的倾向其实标志了对大自然探索的深化.正是庞加莱自己的研究不但不是"为了指出我们祖先推理的错误",反而证明了这种批判性的态度的远见.更令人吃惊的是布朗运动的轨道确与魏尔斯特拉斯的构造的不可求导的连续函数颇为相近!

魏尔斯特拉斯构造的函数是

$$f(x)=\sum_{n=0}^{\infty} b^n\cos(a^n\pi x),\quad -\infty<x<\infty,0<b<1. \tag{1}$$

这里 $a>0$,其值待定.级数(1)显然在 $-\infty<x<+\infty$ 中一致收敛,所以 $f(x)$ 在 $-\infty<x<+\infty$ 上连续.它是无穷多个正弦曲线(余弦函数也是正弦曲线)叠加而成的,记其第 $n+1$ 条曲线为 $y=T_n(x)=b^n\cos(a^n\pi x)$,其周期为 $\frac{2}{a^n}$,振幅为 $b^n$.我们知道,正弦曲线之斜率(按绝对值计)之最大值出现在它的零点处.我们将用这个最大值刻画它陡峭的程度,故不妨称之为 $T_n(x)$ 的陡度,于是 $T_n(x)$ 的陡度就是

$$\max\left|\frac{\mathrm{d}T_n(x)}{\mathrm{d}x}\right|=(ab)^n\pi.$$

现在令

$$a>1\text{ 为一整数,且 } ab>1, \tag{2}$$

所以 $T_{n+1}(x)$ 之陡度又比 $T_n(x)$ 的陡度大 $ab$ 倍,所以构成函数(1)的这些正弦波(即正弦曲线)就越来越窄,越来越陡,其振幅也越来越小.图3-7-1就 $b=\frac{1}{2}$,$a=5$ 画出了级数(1)的三个部分和.短划线是部分和 $S_0(x)=T_0(x)=\cos\pi x$;虚线是部分和 $S_1(x)=T_0(x)+T_1(x)=\cos\pi x+\frac{1}{2}\cos 5\pi x$;实线则是部分和 $S_2(x)=T_0(x)+T_1(x)+T_2(x)=\cos\pi x+\frac{1}{2}\cos 5\pi x+\frac{1}{4}\cos 25\pi x$.图3-7-1给我们的启发是:$T_n(x)$ 的零点与 $T_{n+1}(x)$ 的零点纠合在一起;这就是为什么我们要求 $a$ 为整数的原因.我们还要求 $T_n(x)$ 的极值点(即 $\frac{\mathrm{d}T_n(x)}{\mathrm{d}x}$ 的零点)又与 $T_{n+1}(x)$ 的极值点纠合在一起.因为 $T_n(x)$ 的每个周期内必有 $T_{n+1}(x)$ 的 $a$ 个周期,前者的每个半周中自然含有 $T_{n+1}(x)$ 的 $a$ 个半周.但是 $T_n(x)$ 的最大与最小值交替出现,两个相邻最大与最小值又只相距半周.所以若 $T_n(x)$ 与 $T_{n+1}(x)$ 有一个共同的最大值 $x$,则当 $x$ 从 $T_n(x)$ 的这个最大值走到下一个极值为最小值时,则 $T_{n+1}$ 也由最大变为最小,再变为最大……一共变了 $a$ 次.如果 $a$ 是奇数,则 $x$ 对于 $T_{n+1}(x)$ 而言仍然是由极大变成了极小.正是这样才能保证 $T_n(x)$ 叠

加$T_{n+1}(x)$时是极大叠加到极大上，极小叠加到极小上而不至于互相抵消.我们可以想像：叠加的结果会成为一条“毛茸茸”的曲线，而有可能是一个不可求导的函数的图像.所以我们除(1)以外还假设 $a>1$ 为一奇数. (3)

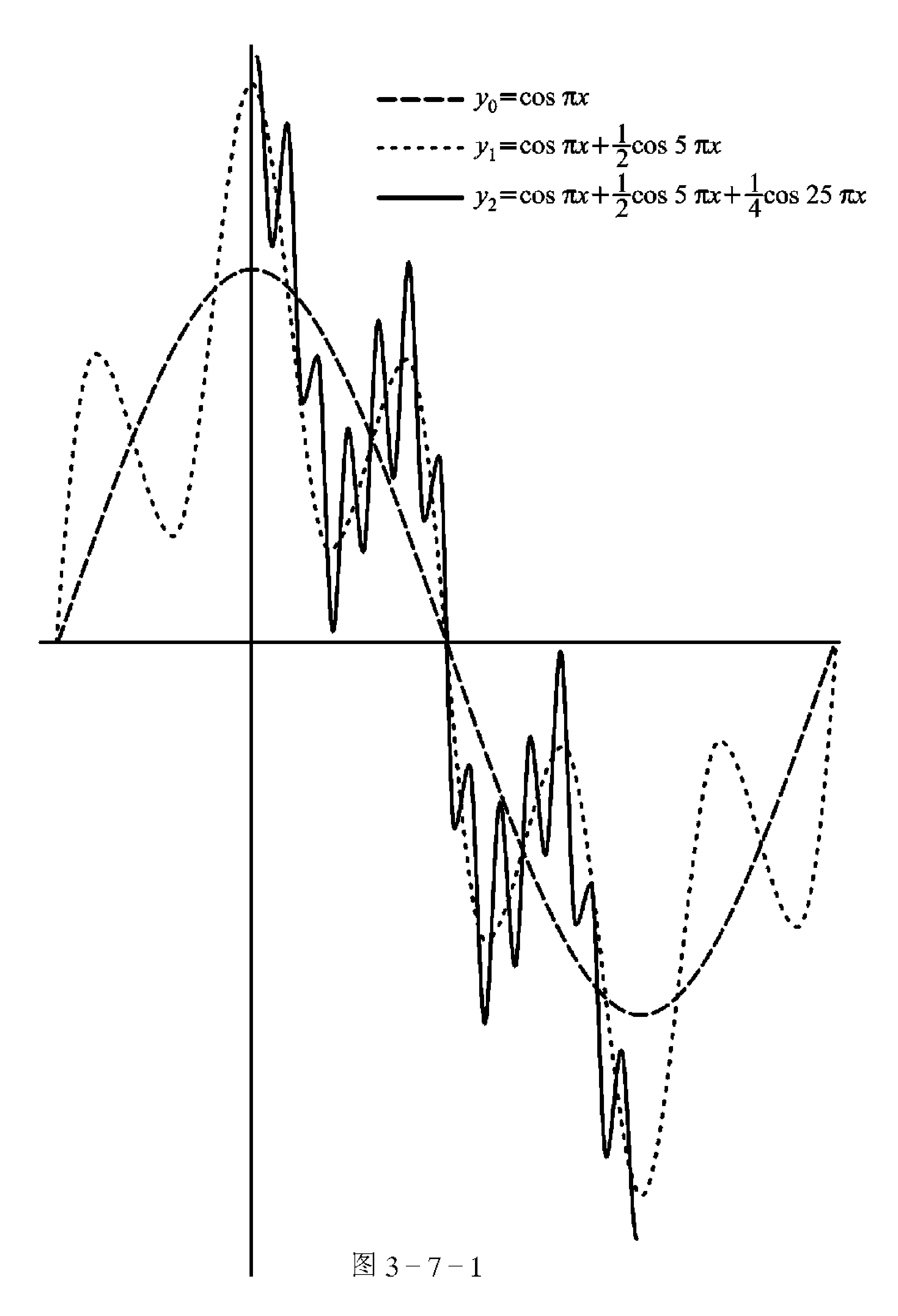



图 3-7-1

下面把这些讨论“数学地”讲明白.

首先看 $T_n(x)$的零点 $x_n^k=\dfrac{2k+1}{2a^n}$.其中 $k$ 是整数.这时 $a^{n+1}x_n^k=\dfrac{2k+1}{2}a$ 也是一个奇数的一半，所以 $T_{n+1}(x_n^k)=b^{n+1}\cdot\cos(a^{n+1}x_n^k\pi)=b^{n+1}\cos\left(\dfrac{2k+1}{2}a\pi\right)=0$.所以 $T_n(x)$的零点一定是 $T_{n+1}(x)$，$T_{n+2}(x)$，…的零点，但不一定是 $T_0(x)$，…，$T_{n-1}(x)$的零点.在函数(1)的图像中 $T_n(x)$的零点总是把 $T_{n+1}(x)$，$T_{n+2}(x)$，…的零点纠合在一起，所以我们称 $x_n^k$ 为 $f(x)$之结点.并以

$$N=\left\{\frac{2k+1}{2a^n},k=0,\pm1,\pm2,\cdots,n=0,1,2,\cdots\right\}$$

记结点之集合，我们看到

**引理 1**　$N$ 是 $\mathbf{R}$ 的一个可数的稠密子集.

再看 $T_n(x)$ 之极值点 $\tilde{x}_n^k=\dfrac{k}{a^n}$，$k$ 是整数，这时 $a^{n+1}\tilde{x}_n^k=ak$ 故 $T_n(\tilde{x}_n^k)=b^n\cos k\pi=(-1)^kb^n$ 是 $T_n(x)$ 之极大（小）值，而 $T_{n+1}(\tilde{x}_n^k)=b^{n+1}\cos ak\pi=(-1)^kb^{n+1}$，故 $\tilde{x}_n^k$ 也使 $T_{n+1}(x)$ 达到极大（小）. 而且与 $T_n(\tilde{x}_n^k)$ 有相同符号，所以极值点也是纠结在一起的，不会互相抵消；又因为各有不同因子 $b^n$，$0<b<1$，求和以后也不会变为正负无穷. 我们称 $\tilde{x}_n^k$ 为 $f(x)$ 的峰点，并以

$$M=\left\{\frac{k}{a^n},k=0,\pm1,\pm2,\cdots,n=0,1,2,\cdots\right\}$$

记峰点之集合，也很容易看到

**引理 2**　$M$ 是 $\mathbf{R}$ 的一个可数稠密子集.

现在来证明函数 (1) 处处没有导数. 为此任取一点 $x_0$ 并记 (1) 在 $x_0$ 之值为 $y_0=f(x_0)$，并且考虑 $T_n(x)$. $x_0$ 必在 $T_n(x)$ 的两个峰点之间；即有一个整数 $\alpha$ 使 $\dfrac{\alpha}{a^n}\leqslant x_0<\dfrac{\alpha+1}{a^n}$. 在这两个峰点中取更接近 $x_0$ 的一个为 $\dfrac{\alpha_n}{a^n}$（若 $x_0$ 是中点，则取左方的一个），则有

$$-\frac{1}{2a^n}<x_0-\frac{\alpha_n}{a^n}\leqslant\frac{1}{2a^n}，或 -\frac{1}{2}<a^nx_0-\alpha_n\leqslant\frac{1}{2}. \tag{4}$$

$\dfrac{\alpha_n}{a^n}$ 也有两个相邻峰点 $x'=\dfrac{\alpha_n-1}{a^n}$，$x''=\dfrac{\alpha_n+1}{a^n}$. 令函数在这两点上之值为 $f(x')=y'$，$f(x'')=y''$，于是得到两个差商

$$\frac{y'-f(x_0)}{x'-x_0},\qquad \frac{y''-f(x_0)}{x''-x_0}. \tag{5}$$

为了证明 $f(x)$ 在 $x_0$ 没有导数，只需证明当 $x',x''\to x_0$ 时，上述两个差商没有相同极限，也不趋向同号的无穷大. 具体说来，我们要证明

**引理 3**　若 $0<b<1$，$a$ 为大于 1 的奇数且

$$ab>1+\frac{3}{2}\pi, \tag{6}$$

则 (5) 中的两个差商不可能有相同的极限，也不趋向同号的无穷大.

**证**　记 $x_{n+1}=x_0a^n-\alpha_n$，由 (4) 有 $|x_{n+1}|\leqslant\dfrac{1}{2}$，而且

$$x'-x_0=\frac{\alpha_n-1-x_0a^n}{a^n}=\frac{1}{a^n}[-1-(x_0a^n-\alpha_n)]=\frac{1}{a^n}(-1-x_{n+1}),$$

$$x''-x_0=\frac{\alpha_n+1-x_0a^n}{a^n}=\frac{1}{a^n}[1-(x_0a^n-\alpha_n)]=\frac{1}{a^n}(1-x_{n+1}).$$

因为 $a>1$，故当 $n$ 充分大时，$|x'-x_0|$ 与 $|x''-x_0|$ 皆可任意小.

再注意到

$$f(x)=\sum_{m=0}^{n-1} b^m\cos(a^m\pi x)+\sum_{m=n}^{\infty} b^m\cos(a^m\pi x)$$
$$=S_{n-1}(x)+R_n(x).$$

所以

$$f(x')=\sum_{m=0}^{n-1} b^m\cos(a^m\pi x')+\sum_{m=n}^{\infty} b^m\cos[(\alpha_n-1)a^{m-n}\pi]$$
$$=\sum_{m=0}^{n-1} b^m\cos(a^m\pi x')-(-1)^{\alpha_n}b^n\sum_{\nu=0}^{n-1} b^\nu,$$
$$f(x'')=\sum_{m=0}^{n-1} b^m\cos(a^m\pi x'')+\sum_{m=n}^{\infty} b^m\cos[(\alpha_n+1)a^{m-n}x]$$
$$=\sum_{m=0}^{n-1} b^m\cos(a^m\pi x'')-(-1)^{\alpha_n}b^n\sum_{\nu=0}^{n-1} b^\nu.$$

注意我们这里利用了 $a^{m-n}$ 为奇数 $2l+1$，故 $\cos[(\alpha_n\pm1)a^{m-n}\pi]=\cos[(\alpha_n\pm1)\pi]=(-1)^{\alpha_n\pm1}=-(-1)^{\alpha_n}$ ($\alpha_n$ 是整数). 同样 $f(x_0)$ 也可以分成两项：利用 $x_{n+1}=x_0a^n-\alpha_n$，而 $x_0=a^{-n}(x_{n+1}+\alpha_n)$，故

$$f(x_0)=\sum_{m=0}^{n-1} b^m\cos(a^m\pi x_0)+\sum_{m=n}^{\infty} b^m\cos[a^{m-n}\pi(x_{n+1}+\alpha_n)]$$
$$=\sum_{m=0}^{n-1} b^m\cos(a^m\pi x_0)+(-1)^{\alpha_n}b^n\sum_{\nu=0}^{\infty} b^\nu\cos(a^\nu\pi x_{n+1}).$$

于是(5)中的两个差商可以写成

$$\frac{y'-f(x_0)}{x'-x_0}=\sum_{m=0}^{n-1} b^m\frac{\cos(a^m\pi x')-\cos(a^m\pi x_0)}{x'-x_0}$$
$$+b^n\sum_{\nu=0}^{\infty} b^\nu(-1)^{\alpha_n+1}\frac{1+\cos(a^\nu\pi x_{n+1})}{x'-x_0},$$
$$\frac{y''-f(x_0)}{x''-x_0}=\sum_{m=0}^{n-1} b^m\frac{\cos(a^m\pi x'')-\cos(a^m\pi x_0)}{x''-x_0}$$
$$+b^n\sum_{\nu=0}^{\infty} b^\nu(-1)^{\alpha_n+1}\frac{1-\cos(a^\nu\pi x_{n+1})}{x'-x_0}.$$

现在分别估计各个差商的两项. 对每一个差商的第一个和式利用和差化积公式有

$$\sum_{m=0}^{n-1} b^m\frac{\cos(a^m\pi x')-\cos(a^m\pi x_0)}{x'-x_0}$$
$$=\sum_{m=0}^{n-1}\frac{b^m}{x'-x_0}2\sin a^m\pi\frac{x'+x_0}{2}\sin a^m\pi\frac{x_0-x'}{2}$$
$$=\sum_{m=0}^{n-1}-a^mb^m\pi\frac{\sin a^m\pi\frac{x_0-x'}{2}}{a^m\pi\frac{x_0-x'}{2}}\sin a^m\pi\frac{x'+x_0}{2}.$$

因为 $\left|\frac{\sin a^m\pi\frac{x_0-x'}{2}}{a^m\pi\frac{x_0-x'}{2}}\right|\leqslant1$，$\left|\sin a^m\pi\frac{x'+x_0}{2}\right|\leqslant1$，考虑到

$a^m b^m \pi > 0$，所以这一和式的绝对值

$$\leqslant \pi \sum_{m=0}^{n-1} a^m b^m = \pi \frac{a^n b^n - 1}{ab - 1} < \frac{\pi a^n b^n}{ab - 1}.$$

因此存在一个常数 $\varepsilon$，$-1 < \varepsilon < 1$，使

$$\text{第一个和式} = \frac{\varepsilon \pi a^n b^n}{ab - 1}. \tag{7}$$

第一个差商的第二个和式是一个收敛的无穷级数．因为 $x' - x_0 = -\frac{1}{a^n}(1 + x_{n+1})$，所以

$$b^n \sum_{\nu=0}^{\infty} b^{\nu} (-1)^{\alpha_n + 1} \frac{1 + \cos(a^{\nu} \pi x_{n+1})}{x' - x_0}$$

$$= (-1)^{\alpha_n} a^n b^n \sum_{\nu=0}^{\infty} b^{\nu} \frac{1 + \cos(a^{\nu} \pi x_{n+1})}{1 + x_{n+1}}.$$

现在新级数第一项即 $\nu = 0$ 的一项是

$$\frac{1 + \cos(\pi x_{n+1})}{1 + x_{n+1}} \geqslant \frac{2}{3}.$$

这是因为 $|x_{n+1}| \leqslant \frac{1}{2}$，从而分子 $1 + \cos(\pi x_{n+1}) \geqslant 1$，而分母 $\leqslant \frac{3}{2}$．总之新级数第一项 $\geqslant \frac{2}{3}$，其余各项则为非负，所以其和 $\geqslant \frac{2}{3}$．记其和为 $\eta' \cdot \frac{2}{3}$，应有 $\eta' \geqslant 1$．合并以上的结果，

$$\text{第一个差商} = \frac{\varepsilon \pi a^n b^n}{ab - 1} + (-1)^{\alpha_n} \frac{2}{3} \eta' a^n b^n.$$

如果把 $\varepsilon$ 写成 $\varepsilon = (-1)^{\alpha_n} \eta' \varepsilon'$，则 $\varepsilon' = \frac{(-1)^{\alpha_n}}{\eta'} \varepsilon$，$|\varepsilon'| = |\varepsilon| / \eta' < 1$，而

$$\text{第一个差商} = (-1)^{\alpha_n} a^n b^n \left(\frac{2}{3} + \frac{\varepsilon' \pi}{ab - 1}\right) \eta' = (-1)^{\alpha_n} a^n b^n p'. \tag{8}$$

这里 $p' = \left(\frac{2}{3} + \frac{\varepsilon' \pi}{ab - 1}\right) \eta' > \frac{2}{3} - \frac{\pi}{ab - 1} > 0$．这里我们应用了(6)式．

第二个差商可以类似地处理，注意到这时 $x'' - x_0 = \frac{1}{a^n}(1 - x_{n+1})$，我们将得到

$$\text{第二个差商} = (-1)^{\alpha_n + 1} a^n b^n \left(\frac{2}{3} + \frac{\varepsilon'' \pi}{ab - 1}\right) \eta'' = (-1)^{\alpha_n + 1} a^n b^n p''. \tag{9}$$

这里 $p'' = \left(\frac{2}{3} + \frac{\varepsilon'' \pi}{ab - 1}\right) \eta'' > \frac{2}{3} - \frac{\pi}{ab - 1} > 0$.

总结以上即知：

若 $(-1)^{\alpha_n} = 1$，有

$$\frac{y' - f(x_0)}{x' - x_0} \geqslant a^n b^n \left(\frac{2}{3} - \frac{\pi}{ab - 1}\right),$$

$$\frac{y'' - f(x_0)}{x'' - x_0} \leqslant -a^n b^n \left(\frac{2}{3} - \frac{\pi}{ab - 1}\right).$$

当 $(-1)^{\alpha_n} = -1$ 时

$$\frac{y' - f(x_0)}{x' - x_0} \leqslant -a^n b^n \left(\frac{2}{3} - \frac{\pi}{ab-1}\right),$$

$$\frac{y'' - f(x_0)}{x'' - x_0} \geqslant a^n b^n \left(\frac{2}{3} - \frac{\pi}{ab-1}\right).$$

由于 $x'$，$x''$属于 $\mathbf{R}$ 上的一个可数稠密集 $M$，故知函数(1)的差商不可能有极限，也不可能趋向于同符号的无穷大. 魏尔斯特拉斯的函数(1)在任一点 $x_0$ 都没有导数得证.

# 第四章　积　分　学

## §1　这样评论黎曼公正吗？

**1．不连续函数进入了人们的视野**　迪厄多内（Jean Dieudonné）在《现代分析基础》（Éléments d' Analyse）第一卷（中译本由科学出版社 1982 年出版）159 页说了这样一段话，以解释为什么这部大书（全书共九卷）中完全不讲黎曼积分："如果不是它的有权威的名字，它老早就该没落下去了．因为对任何一位从事研究工作的数学家来说（带着对黎曼的天才应有的尊敬），十分清楚，现今这一"理论"的重要性在测度与积分的一般理论中，最多不过是一普通的有趣的练习（参看13.9 问题 7），只有那种学究传统的顽固的保守主义才会把它冻结成课程的正规部分，长时间以后必将失去它的历史重要性．"现在我们要追溯一下历史，看一看黎曼是在什么样的历史背景下提出他的积分理论，以及他的理论的不足之处究竟在什么地方．

这就需要从傅里叶讲起．傅里叶从研究大量物理问题（主要是热传导问题）出发，指出"任何函数"$f(x)$，$x\in[-\pi,\pi]$都可以写成一个无穷级数

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n\cos nx+b_n\sin nx),\tag{1}$$

这里

$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(\xi)\cos n\xi\mathrm{d}\xi,b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(\xi)\sin n\xi\,\mathrm{d}\xi.\tag{2}$$

在傅里叶（1768—1830）的时代（19 世纪初），什么是"任何函数"是一个没有说清的问题，本书第二章讲到了这件事．当时，一种说法是：函数就是一个解析表达式．因此，不同的解析表达式就表示不同的函数；另一种说法则是：顺手一笔画出来的曲线就是一个函数．这是同一个人欧拉，在不同场合下讲的话．傅里叶在他的名著《热的解析理论》（Jean-Baptiste Joseph Fourier，Théorie Analytique de la Chaleur，中译本武汉出版社 1993 年出版）中讨论三角函数的一般理论时（中译本 132～133 页），就给出了一个例子（见图 4－1－1）

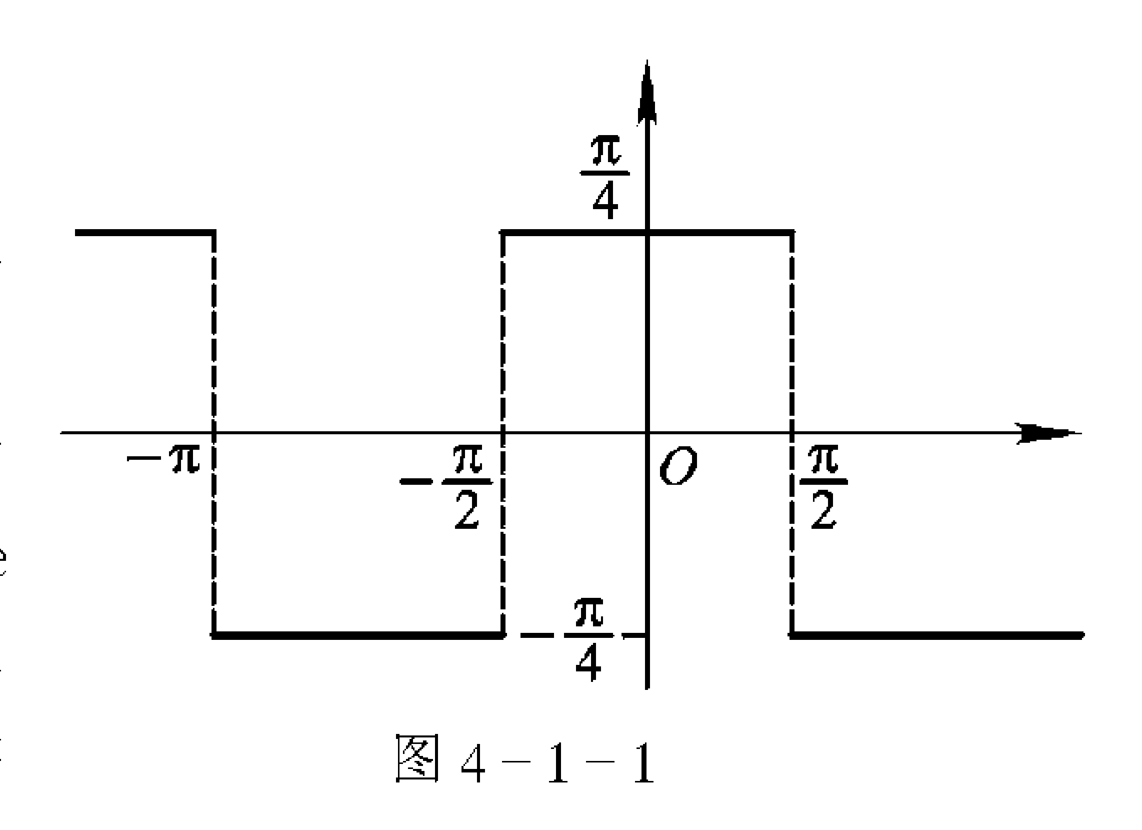


图 4－1－1

$$\cos-\frac{1}{3}\cos 3x+\frac{1}{5}\cos 5x-\ =\begin{cases}\frac{\pi}{4},|x|<\frac{\pi}{2},\\-\frac{\pi}{4},\frac{\pi}{2}<|x|<\pi.\end{cases}\tag{3}$$

同一个级数怎么能表示不同的函数呢？看来并不是所有的函数都如雅可比说的是"合理的"函

数，而可以是不连续的，而且研究这种有不连续点的函数，在物理上也是“合理的”.这一个例子以及后来越来越多的例子才引导到关于一般函数的“狄利克雷定义”，即本书第二章所介绍的定义.越来越多的例子表明，函数的不连续点的集合构造可以是十分复杂的.例如“狄利克雷函数”

$$\chi(x)=\begin{cases}1, x\text{ 为有理数}, \\ 0, x\text{ 为无理数}\end{cases} \tag{4}$$

是处处不连续的，而形状十分相近的“黎曼函数”

$$R(x)=\begin{cases}1/q, x=\dfrac{p}{q}\text{为既约有理数}, \\ 0, \quad x\text{ 为无理数}.\end{cases} \tag{5}$$

却只在有理点不连续.其实傅里叶所说的“任何函数”是指的连续函数.但是如(3)那样的函数在傅里叶看来也是连续的，因为它的图像如图 4－1－1 所示是一条连续曲线，所以表示连续函数.傅里叶把图 1 中的虚线也包括到(3)的图像中去了，而按狄利克雷的定义这样做是不许可的，因为这样一来在例如 $x=\dfrac{\pi}{2}$ 处，函数的值就不确定了.再用柯西关于连续函数的定义才看出这个点是不连续点.不过傅里叶的书实际上成于 1810 左右，狄利克雷的函数定义和柯西关于连续性的定义乃是 1820—1830 左右的事，何况他们二人自己有时也会“糊涂”的.不论如何，当函数的不连续性逼迫人们予以正视时，从定义傅里叶系数 $a_n$ 与 $b_n$ 的(2)式看，有必要研究不连续函数的积分.

把定积分定义为积分和的极限

$$\int_a^b f(x)\mathrm{d}x=\lim\sum f(\xi_i)(x_i-x_{i-1}) \tag{6}$$

的是柯西.但是他假设了 $f(x)$在$[a,b]$上连续.如果 $f(x)$在 $x=c$ 时有间断，则定义

$$\int_a^b f(x)\mathrm{d}x=\int_a^c f(x)\mathrm{d}x+\int_c^b f(x)\mathrm{d}x.$$

现在用的反常积分的定义也是柯西给出的.但是比较一般的函数 $f(x)$的积分如何定义，则是黎曼 1854 年在研究三角级数(即形如(1)之右方的级数，但是并不确定是否有一个 $f(x)$存在使(2)成立)时提出来的，当时黎曼已意识到，这一项研究虽然不一定有直接的物理学上的应用，在数学上却是重要的.

上面我们看见了两个特例(4)与(5)，这使人产生一个疑问：为什么要研究这些人为的特例？上一章§7 中我们引用了庞加莱的话，然后又引用了佩韩关于布朗运动的议论.那么，研究这些古怪的函数是为了证明自己前辈的无能吗？难道黎曼(还有魏尔斯特拉斯)还需要靠“揭发”自己前辈的无能才能巩固自己的地位吗？或者是不是这些函数与某些物理现象有关？都不是，上一章讲到佩韩关于处处不可微函数与布朗运动的关系的话，当时只是一种设想而已.有一种偏见，以为数学的发展是一帆风顺的：从几个不辩自明的真理(公理、公设)，以及任意设定的某些对象的性质(定义)出发，按照一些确定的推理规则，就可以得出一个又一个确定无疑的真理.从这个意义上说，数学的方法论是“先验”的.上一章§6 中我们也说过这个意思.但是数学之为“先验”只是相对于其它科学之“后验”而言，上一章我们正是相对于力学中最小作用的原理的演变讲数学方法论的先验性质的.其实，数学的发展同样充满了争论和矛盾.本书各章都说明这个问题.前辈数学家发现了一些真理，以为它是正确的：后来的数学家发现了其中有不少毛病、漏洞，而且这些毛病和漏洞常以反例形式出现.这样，人们去清理前辈的工作，补足他们的缺失，发展他们的成

就.甚至许多定义也不是确定于证明之前，而是产生了证明过程之中.举一个例，早在莱布尼茨就认为如果连续函数的级数 $\sum_{n=1}^{\infty} u_n(x)$ 收敛，则其和也一定连续.柯西也是这样想的，而且(怀着对莱布尼茨天才应有的尊敬)力图去证明它，甚至傅里叶的例子(3)开始也不被认为是反例，因为傅里叶对和函数的图形补上了一些虚线，使之成为“连续曲线”.直到柯西给出了连续函数的定义后，才知道这个连续曲线并非连续函数的图像.但是柯西仍然力图去证明那个可以追溯到莱布尼茨的“定理”，他的这种努力甚至见于他为数学分析严格性奠基的力作《代数分析教程》(Cours d'Analyse Algebrique，1821).直到 1876 年，阿贝尔指出，柯西的“定理”有例外.(定理怎么能有例外？)最终打破这个哑谜的是塞德尔(P.L.Seidel).他在 1847 年分析柯西用 $\varepsilon$—$N$ 语言来表达的证明时指出，这个 $N$ 不仅依赖于 $\varepsilon$，还依赖于 $x$：$N=N(\varepsilon,x)$.如果 $x$ 只能取有限多个值，从这些 $N(\varepsilon,x)$ 中自然可以找出最大的一个作 $N(\varepsilon)$.但若 $x$ 可以取无穷多个值(例如某一区间 $[a,b]$ 中的一切值)，则 $N(\varepsilon,x)$ 就不一定有有限的最大值 $N(\varepsilon)$ 以适合一切 $x$ 点了，因此柯西的证明失效.赛德尔在自己的论文中说了一段话：“方才已经确认，定理不是普遍有效的，因为它的证明必定依赖某个额外的隐蔽假定(即 $N(\varepsilon,x)$ 对一切 $x$ 有有限的最大值——本书作者).有鉴于此，我们要给证明作一次更细微的分析.发现那个隐蔽的假设倒不很难.于是可以反推这个假设所表达的条件确是表示不连续函数的级数满足不了的.因为，只有这样，那串别无他错的证明才跟另一头的确凿事实才能重新归于一致.”这就是说，从分析柯西的证明入手，人们找出了漏洞，发现了收敛的级数应分为两类，一类是 $N(\varepsilon,x)$ 有最大值的，另一类则是没有的.至此，一致收敛性的定义呼之欲出：并不是人们先验地有了一致收敛性定义，并由之证明了相应定理，而是从证明过程中这个定义自己跑到我们面前.我们只好承认它，应用它.当时看到了这一点的不只有赛德尔，还有斯托克斯.魏尔斯特拉斯稍早一些也在幂级数研究中独立地得到了一致收敛性概念.到底是谁首先明确宣布这个定义，现在人们一般地归功于魏尔斯特拉斯.不过斯托克斯，赛德尔还有傅里叶本人都指出它与收敛速度有关.(以上材料引自拉卡托斯《证明与反驳》，附录一，中译本 152～170 页).一致收敛的案例说明了数学的发展并不是先验的，反例的重要性在于它揭露了无法回避的矛盾.数学的发展并不是为了说明种种反例，相反，研究反例是为了发展数学.19 世纪最后二三十年出现了大量“反例”，正表明数学的深入发展，在数学对世界的认识的过程中起了极大作用.这可以说是现代数学的一大特点.黎曼积分正是这种历史背景下出现的第一个系统的积分理论.回忆到上一章讲的处处不可微函数的例子是魏尔斯特拉斯 1861 年发现的，可以看到，不连续函数进入人们的视野是时代的产物.

**2．函数的可积性**　黎曼第一个提出可积性问题，即要求刻画出使(6)之极限存在的函数类.(这里说明一下，本节中讲到的 $f(x)$ 都是实值函数.若 $f(x)$ 是复值的，则划分开实虚部处理).人们通常都说“积分是积分和的极限”，这个说法并不准确，我们学过了序列的极限与函数的极限，但(6)式右方的积分和

$$\sum_{i=1}^{n} f(\xi_k)(x_i-x_{x-1}) \tag{7}$$

既不构成序列，也不是某连续变量的函数.为了明确地了解它，我们首先要求 $f(x)$ 定义在有限区间 $[a,b]$ 上，而且 $f(x)$ 不可以无界.对 $[a,b]$ 作一个分划 $P$：$a=x_0<x_1$

上(下)确界 $M_i\,(m_i)$. $-\infty < m_i \leqslant M_i < +\infty$. 反之,若 $f(x)$无界,$M_i$, $m_i$ 中必有 $\pm\infty$,这会使下式没有意义. 我们不考虑 $f(x)$在$[x_{x-1}, x_i]$中是否可以达到 $m_i$ 与 $M_i$,而定义 $f(x)$相应于分划 $P$ 的上和 $S_P$ 与下和 $s_P$:

$$S_P = \sum_{i=1}^{n} M_i\,(x_i - x_{i-1}),$$

$$s_P = \sum_{i=1}^{n} m_i\,(x_i - x_{i-1}). \tag{8}$$

很清楚

$$s_P \leqslant S_P. \tag{9}$$

但对不同的分划 $P'$与 $P''$, $s_{P'}$与 $S_{P''}$可否比较? 首先我们注意到,若在分划 $P$ 中加一个分点 $x'$(称为 $P$ 的"加细",加细后的新分划为 $P'$)例如,$x_{i-1} < x' < x_i$,则 $s_P$ 和 $S_P$ 变成

$$s_{P'} = \sum_{j=1}^{i-1} m_j\,(x_j - x_{j-1}) + m_i'\,(x' - x_{i-1}) + m_i''\,(x_i - x') + \sum_{j=i+1}^{n} m_j\,(x_j - x_{j-1}),$$

$$S_{P'} = \sum_{j=1}^{i-1} M_j\,(x_j - x_{j-1}) + M_j'\,(x' - x_{i-1}) + M_i''\,(x_i - x') + \sum_{j=i+1}^{n} M_j\,(x_j - x_{j-1}).$$

这里 $m_i' = \inf\limits_{[x_{i-1}, x']} f(x)$, $m_i'' = \inf\limits_{[x', x_i]} f(x)$, $M_i'$, $M_i''$ 意义自明. 于是 $m_i = \min(m_i', m_i'')$, $M_i = \max(M_i', M_i'')$,而可见 $s_P \leqslant s_{P'}$, $S_{P'} \leqslant S_P$,就是说,分划加细后下和不减而上和不增. 把 $P'$与 $P''$合并成一新分划$P$(不妨写作 $P' + P'' = P$),则 $P$ 同时是$P'$与$P''$的加细. 因此

$$s_{P'} \leqslant s_P \leqslant S_P \leqslant S_{P''} \tag{10}$$

就是说:不论哪一个分划的下和一定不大于任一分划的上和. 于是若固定一分划 $P_0$,可见对一切分划 $P$, $s_P \leqslant S_{P_0}$. 从而$\{s_P\}$有界而有上确界,记之为

$$\underline{I} = \sup_P \{s_P\},$$

同理,$\{S_P\}$有下界而有下确界 $\overline{I}$:

$$\overline{I} = \inf_P \{S_P\}.$$

而且

$$\underline{I} \leqslant \overline{I}.$$

$\underline{I}$和 $\overline{I}$分别称为 $f(x)$在$[a, b]$上的下积分与上积分. 它们是某集合的上、下确界,而不是序列的极限. 看下面的定义即知,说积分是积分和的极限是不准确的.

**定义 1**　若 $\underline{I} = \overline{I}$,则称 $f(x)$在$[a, b]$上黎曼可积($R$－可积,在不产生误解时就说可积),其公共值 $I = \underline{I} = \overline{I}$称为 $f(x)$在$[a, b]$上的定积分,记作

$$I = \int_a^b f(x)\,\mathrm{d}x \tag{11}$$

以上是达布(G. Darboux)对黎曼的讲法作的一个修正. 黎曼本人仍是使用的积分和(7). (7)与 $I$ 的关系以后再讲,目前我们先用达布上下和来讨论 $f(x)$可积的充分必要条件.

**定理1** $[a,b]$上的有界函数 $f(x)$为可积的充分必要条件是：对任一 $\varepsilon>0$ 均可找到一个分划 $P$ 使

$$0\leqslant S_P-s_P<\varepsilon.$$

**证** **必要性**. 设 $f(x)$可积，则 $\underline{I}=\overline{I}=I$. 由上、下确界的定义，必可找到分划 $P'$和$P''$使

$$s_{P'}>\underline{I}-\frac{\varepsilon}{2}=I-\frac{\varepsilon}{2},$$

$$S_{P''}<\overline{I}+\frac{\varepsilon}{2}=I+\frac{\varepsilon}{2}.$$

令 $P=P'+P''$，则

$$0\leqslant S_P-s_P\leqslant S_{P''}-s_{P'}<I+\frac{\varepsilon}{2}-I+\frac{\varepsilon}{2}=\varepsilon.$$

**充分性**. 由上下确界的定义 $\overline{I}\leqslant S_P$，$\underline{I}\geqslant s_P$，但 $\overline{I}\geqslant\underline{I}$，所以

$$0\leqslant\overline{I}-\underline{I}\leqslant S_P-s_P<\varepsilon.$$

由 $\varepsilon$ 之任意性有 $\overline{I}=\underline{I}$，即 $f(x)$可积，定理证毕.

注意到

$$S_P-s_P=\sum_{i=1}^{n}(M_i-m_i)(x_i-x_{i-1}).$$

为使 $S_P-s_P$ 充分小，就应该要求在相当多的小区间$[x_{x-1},x_i]$上 $M_i-m_i$ 充分小，而使 $M_i-m_i$ 不太小的小区间$[x_{i-1},x_i]$之总长很小. 但 $M_i-m_i$ 相当小就很接近于函数的连续性. 所以为使 $S_P-s_P$ 可以任意小，就要求在相当多的小区间上 $f(x)$连续，而 $f(x)$不连续的点不太多. 这样就把 $f(x)$之连续性与可积性联系起来了. 这一段话很不明确，读者完全可以认为是说错了. 为了把这一点弄明确，我们引进函数振幅的概念. 现在我们有意地设 $f(x)$是 $n$ 元函数. 设 $f:A\subset\mathbf{R}^n\to\mathbf{R}$，$a\in A$，$f(x)$在 $a$ 点不连续的程度可以度量如下：令

$$M(f,a,\delta)=\sup\{f(x):x\in A\text{ 且 }|x-a|<\delta\},$$
$$m(f,a,\delta)=\inf\{f(x):x\in A\text{ 且 }|x-a|<\delta\}.$$

我们称 $O(f,a)=\lim\limits_{\delta\to0}[M(f,a,\delta)-m(f,a,\delta)]$为 $f$ 在 $a$ 点的振幅. 由于 $M(f,a,\delta)-m(f,a,\delta)$是非负的随 $\delta$ 不增的函数，所以 $O(f,a)$存在. 显然 $f$ 在$a$ 点相对于$A$ 连续(不连续)即 $O(f,a)=0(\neq0)$，这里我们特意加上"相对于 $A$"几个字，因为定义 $M$ 与$m$ 时并没有说$x$ 在球$|x-a|<\delta$ 内，而是在 $A\cap\{|x-a|<\delta\}$内(注意，我们的讨论不只对一元函数 $f(x)$适用，而对 $x=(x_1,\cdots,x_n)$时的多元函数 $f(x)$也适用. 这时$|x-a|<\delta$ 确实是球体而不是区间.)即只用到此球含于 $A$ 内的部分，因为 $f(x)$只定义于 $A$ 上，而不是定义在整个球$|x-a|<\delta$ 上(当然也不会对此球内一切 $x$ 点$f(x)$均无定义：至少对中心 $x=a$，$f(x)$是有定义的)，这样得到的连续性是相对于 $A$ 的连续性. 在一元函数情况下，若 $f(x)$定义在 $A=[a,b]$上，则在定义域的左右两端只讨论 $f(x)$的单侧连续性，这就是相对于$[a,b]$的连续性. 但在多元函数情况，由于定义域 $A$ 的构造可以十分复杂，相对于 $A$ 的连续性也复杂多了. 例如，若 $a\in A$ 是一个孤立点，则$\{x\in A,|x-a|<\delta\}$当 $\delta$ 充分小时就只有一个$a$ 点，这时自然有 $O(f,a)=0$，即 $f$ 在$a$ 点相对

于$A$连续.由于这个原因,在研究多元函数时,我们常限于$f(x)$定义于一开集上.由于开集中的点都是内点,就不会出现上面的麻烦.总之,多元函数在边界点上的情况是要特别细心处理的.这一类问题在第六章讲到子空间拓扑时还会详细说.

**引理 2**　设有界函数$f(x)$定义在$[a,b]$上,若对一切$x\in[a,b]$均有$O(f,x)<\varepsilon$,则必存在一个分划$P$,使

$$0\leqslant S_P-s_P<2\varepsilon(b-a).$$

**证**　对一切$x\in[a,b]$均存在一个开区间$[x-\delta,x+\delta]$使$M(f,x,\delta)-m(f,x,\delta)<2\varepsilon$.开区间族$\{(x-\delta,x+\delta)\}$覆盖$[a,b]$,故可由其中选出有限多个$U_1,\cdots,U_m$使$[a,b]\subset\bigcup\limits_{k=1}^{m}U_k$,适当选分划$P$可使$P$之每一个小区间$[x_{i-1},x_i]$均含于一个$U_k$中,于是$M_i-m_i<2\varepsilon$,而

$$0<S_P-s_P=\sum_{i=1}^{n}(M_i-m_i)(x_i-x_{i-1})<2\varepsilon(b-a).$$

下面我们就可以联系着$f(x)$的连续性来讨论$f(x)$可积的充分必要条件了.我们的结论是$f(x)$可积的充分必要条件为其不连续点构成一个0测度集(准确些说是0勒贝格(Lebesgue)测度或$0L$测度集):似乎有必要先讲什么是测度(勒贝格测度、$L$测度),但是事实上不需要.介绍这个概念我们又是对$n$维情况来讲.

**定义 2**　设$A\subset\mathbf{R}^n$,我们说$A$具有0测度即指对任一$\varepsilon>0$均可找到可数多个平行于坐标轴的闭长方体$\{U_1,\cdots,U_m,\cdots\}$使$A\subset\bigcup\limits_{k=1}^{\infty}U_k$,而且$\sum\limits_{k=1}^{\infty}\mathrm{vol}(U_k)<\varepsilon$,$\mathrm{vol}(U_k)$即指$U_k$之体积.

**注 1**　定义中的$U_k$也可取为开长方体,只要把每个闭的$U_k$代以同心、平行而稍大的开长方体即可.

**注 2**　任意可数点集$\{x_k\}$必有0测度,因为我们可以把$x_k$放在一个体积为$\frac{\varepsilon}{2^{k+1}}$的闭长方体的中心,而这样得到的可数多个闭长方体之总体积是$\varepsilon\sum\limits_{k=1}^{\infty}\frac{1}{2^{k+1}}=\varepsilon$是任意小数.特别是$[0,1]$中的有理数构成一个零测度集.

**注 3**　有限多个$\{U_1,\cdots,U_m\}$也可以算作可数多个$\{U_1,\cdots,U_m,\cdots\}$,只要假设$U_{m+k}=\varnothing$,$k=1,2,\cdots$即可,但此定义中必须说可数多个$U_k$,若只用有限多个$U_k$将得到另一概念.

**定义 3**　设$A\subset\mathbf{R}^n$,我们说$A$具有0容度(0若尔当(Jordan)测度,$0J$测度)即是对任一$\varepsilon>$

0

$B_\varepsilon \subset [a,b]$，所以在这个开覆盖中能取有限多个$\{U_1,\cdots,U_m\}$覆盖$B_\varepsilon$，且$\sum_{k=1}^{m}\mathrm{vol}(U_k)<\varepsilon$. 把每一个$U_k$换成$\overline{U}_k$仍可覆盖$B_\varepsilon$且总长度不变，所以不妨设这些$U_k$均为闭区间.

对于任一个分划$P$，有

$$S_P - s_P = \sum_{i=1}^{m}(M_i - m_i)(x_i - x_{i-1}).$$

正如前面已指出的，为使$S_P - s_P$充分小，我们或者要求$M_i - m_i$对大多数子区间充分小，或者要求使$M_i - m_i$不太小的区间总长度很小，所以我们把$P$中的小区间分成两组：

$S_1$：与$B_\varepsilon$不相交的小区间.

$S_2$：完全含于某个$U_k$内的小区间.

若一个小区间既不全含于任一$U_k$内又与$B_\varepsilon$相交，则可将它细分为两部分，一部分与$B_\varepsilon$不相交，另一部分含于某个$U_k$内. 所以将$P$的小区间分为两组是可能的.

对$S_1$中的小区间，因$O(f,x)<\varepsilon$，故由引理2，必要时对$P$加细为$P'$即可知，对$P'$中$S_1$类小区间有

$$\sum(M_i - m_i)(x_i - x_{i-1})<2\varepsilon(b-a).$$

对$S_2$中的小区间，因为$f(x)$在$[a,b]$上有界，故$(M_i - m_i)\leqslant 2\sup\limits_{[a,b]}|f(x)| = 2M$，而对这些小区间

$$\sum(M_i - m_i)(x_i - x_{i-1})<2M\sum_{k=1}^{m}\mathrm{vol}|U_k|<2M\varepsilon.$$

把这两个式子加起来即有

$$0\leqslant S_{P'} - s_{P'}<2\varepsilon(b-a)+2M\varepsilon.$$

因此，$f(x)$在$[a,b]$上可积.

反过来设$f(x)$可积，今证$B$有0测度. 但是$B = B_1\cup B_{\frac{1}{2}}\cup\cdots\cup B_{\frac{1}{m}}\cup\cdots$，所以今证每个$B_{\frac{1}{m}}$均有0测度即可. 任给$\varepsilon>0$，因已设$f(x)$可积，故必有一个分划$P$使$S_P - s_P<\varepsilon/m$，用上面同样的方法将$P$中的小区间分成两类，我们只看与$B_{\frac{1}{m}}$相交的那些小区间$[x_{i-1},x_i]$. 若$P$相当细，则在其上$M_i - m_i>\dfrac{2}{m}$，把这些小区间合并起来，有

$$\frac{2}{m}\sum\mathrm{vol}(U_k)\leqslant\sum(M_i - m_i)(x_i - x_{i-1}) = S_P - s_P<\frac{\varepsilon}{m}.$$

所以$\sum\mathrm{vol}(U_k)<\dfrac{\varepsilon}{2}$. 由$\varepsilon$之任意性可知$B_{\frac{1}{m}}$为0容度的. 把可数多个0容度集并起来自然可得0测度集. 所以$B$为0测度集而$f(x)$几乎处处连续，定理证毕.

至此为止我们都是应用达布的上、下和. 怎样转到黎曼所用的积分和(7)呢？

**定理4** 设$f(x)$在$[a,b]$上有界，对任一$\varepsilon>0$，均存在$\delta>0$，使对任何分划$P$，只要$\delta_P<\delta$，就有$s_P>\underline{I}-\varepsilon$，$S_P<\overline{I}+\varepsilon$.

**证** 不失一般性可设$f(x)\geqslant 0$(必要时用$f(x)+C$代替$f(x)$，$C$是充分大的正数)，由$\overline{I}$

之定义,必有分划 $P_0: a=x_0<x_1<\cdots<x_n=b$,使 $S_{P_0}<\bar{I}+\frac{\varepsilon}{2}$,设 $M=\sup\limits_{[a,b]}f(x)$,今证 $\delta=\varepsilon/4(n+1)M$ 即合于所求.

设 $P$ 是任一分划,和证明定理 3 一样,把 $P$ 的小区间分为两组:一组是完全含于 $[x_i-\delta, x_i+\delta]$的,其余列入第二组.于是 $S_P$ 也分成了两项 $S_P^{(1)}$ 和 $S_P^{(2)}$,$S_P^{(1)}$ 中最多有 $n+1$ 项,每一项都含了 $P_0$ 的一个分点($P_0$ 一共有 $n+1$ 个分点),所以这些项相应于 $P$ 中的 $n+1$ 个小区间,总长为 $2(n+1)\delta=\frac{\varepsilon}{2M}$,第二组的小区间由于规定其中不能含有任一个分点,而且其长均不大于 $\delta$,所以必全位于某个小区间$[x_{x-i}, x_i]$内,因此

$$S_P=S_P^{(1)}+S_P^{(1)}\leqslant M\cdot\frac{\varepsilon}{2M}+\sum_{i=1}^{n}M_i(x_i-x_{i-1})<\frac{\varepsilon}{2}+\bar{I}+\frac{\varepsilon}{2}=\bar{I}+\varepsilon.$$

同理 $s_P>\underline{I}-\varepsilon$,引理证毕.

**定理 5** 设 $f(x)$在$[a,b]$上有界并可积,则对任一 $\varepsilon>0$,必可找到 $\delta>0$ 使若某一分划 $P$ 的$\delta_P<\delta$,不论如何选(7)式中的 $\xi_k(x_{k-1}\leqslant\xi_k\leqslant x_k)$都有

$$\left|\sum_{i=1}^{n}f(\xi_i)(x_i-x_{x-i})-\int_a^b f(x)\mathrm{d}x\right|<\varepsilon.$$

其逆亦真.

**证** 因为 $m_i\leqslant f(\xi_i)\leqslant M_i$,所以

$$s_P\leqslant\sum_{i=1}^{n}f(\xi_i)(x_i-x_{i-1})\leqslant S_P.$$

另一方面,由于已设 $f(x)$可积,故 $\underline{I}=\bar{I}=I$,而定理 4 又推出,只要 $\delta_P<\delta$,必有 $s_P>\underline{I}-\frac{\varepsilon}{2}$,$S_P<\bar{I}+\varepsilon$ 所以当$\delta_P<\delta$ 时

$$I-\varepsilon=\underline{I}-\varepsilon<s_P\leqslant\sum_{i=1}^{n}f(\xi_i)(x_i-x_{i-1})\leqslant S_P<\bar{I}+\varepsilon=I+\varepsilon,$$

而定理得证(逆定理是明显的).

这样我们看到,不论是用黎曼的积分和或达布的上下和来定义定积分,结果都是一样的.我们也由此看到了所谓定积分是积分和的极限这句话的确切含意.

由定理 5 我们就自然会提出,究竟哪一些定义于$[a,b]$上的函数是可积的?当然连续函数,只有有限多个不连续点的函数,乃至有可数多个不连续点的函数都是可积的.最后这种情况是由于可数点集均为 0—测度集,证如下:令不连续点为 $x_1<x_2<\cdots<x_n<\cdots$,作闭区间 $U_k=[x_k-\frac{\varepsilon}{2^k}, x_k-\frac{\varepsilon}{2^k}]$,显然$\{x_1,x_2,\cdots,x_n,\cdots\}\subset\bigcup\limits_{k=1}^{\infty}U_k$(注意 $U_k\cap U_l$ 可以是非空的),而 $\sum\limits_{k=1}^{\infty}\mathrm{vol}(U_k)=2\sum\limits_{k=1}^{\infty}\varepsilon/2^k=2\varepsilon$.证毕.具有可数多不连续点的函数的一个重要情况是单调函数 $f(x)$.这是因为 $\inf\limits_{[a,b]}f(x)=f(a)$,$\sup\limits_{[a,b]}f(x)=f(b)$(设 $f(x)$为不减的).记 $M=f(b)-f(a)$,$B=\{x: f$ 在 $x$ 点不连续$\}$,因为单调不减函数的每个不连续点 $x$ 都是第一类间断点,而 $O(f,x)=f(x+0)-$

$f(x-0)$，仿照前面的记号 $B=\bigcup_{m=1}^{\infty}B_m$，$B_m=\{x: 0(f,x)\geqslant\frac{M}{m}\}$，则 $B_m$ 必为有限集.因若不然，$B_m$ 中至少有可数多个点 $x_1<x_2<\cdots$ 而 $f(b)-f(a)\geqslant\sum_{i=1}^{\infty}[f(x_i+0)-f(x_i-0)]\geqslant\sum_{i=1}^{\infty}\frac{M}{m}=+\infty$. $B$ 作为可数多个有限集之并当然最多是可数的.

**推论 6** $[a,b]$上的单调函数必为可积的.

**3. 微积分的基本定理** 以上我们系统地讨论了函数可积性问题，下面转向另一个问题：积分和微分的关系问题.人们心目中都以为二者关系很简单，即二者互为逆运算.认识到二者互为逆运算十分重要，因为微分表反过来看就是积分表.自此以后，人们就知道积分是怎样计算的.如果没有这样一个表，遇到任何一个问题都得利用第一章里介绍的希腊人的穷竭法去处理，微积分就决不可能成为研究种种实际问题的工具，不可能为物理学家与工程师所掌握.所以牛顿和莱布尼茨把他们“发现”的这个公式

$$\int_a^b F'(x)\mathrm{d}x=F(b)-F(a) \tag{12}$$

称为微积分的基本定理.但是在牛顿和莱布尼茨的时代，对数学分析的基本概念还很不清楚，所以他们的论证也很不清楚.直到柯西、黎曼的时代，这个定理才有明确的证明.柯西对连续的 $F'(x)$ 证明了(12)，达布则对一般可积的 $F'(x)$ 证明了它.

我们首先直观地分析一下(12)的含意.因为微分和积分是如此不同的概念，而有如(12)这样的公式将它们联结起来当然是很有意思的.为此，我们先把“极限”的成分排除掉，而把“连续”的东西代以相应的“离散”的东西如下：

$$\int_a^b \quad 代以 \quad \sum_{i=1}^{n}$$

$$f'(x) \quad 代以 \quad \left(\frac{\Delta f}{\Delta x}\right)_i=\frac{f(x_i)-f(x_{i-1})}{x_i-x_{i-1}}$$

$$\mathrm{d}x \quad 代以 \quad (\Delta x)_i=x_i-x_{i-1}$$

这样一来，(12)就化为一个几乎自明的公式

$$\sum_{i=1}^{n}\frac{\Delta f}{\Delta x_i}(\Delta_x)_i=\sum_{i=1}^{n}(\Delta f)_i=\sum_{i=1}^{n}[f(x_i)-f(x_{i-1})]$$
$$=f(x_n)-f(x_0)=f(b)-f(a).$$

可见，(12)原来是一个很简单的组合学公式.组合学成分是积分的重要因素，在本书最后一章我们还要讨论这一件事.但现在我们把极限成分放进去，这时就会发现(12)中其实包含了两个不同的问题.

(1) 若 $f(x)$ 可积，则它的不定积分 $\int_a^x f(\xi)\mathrm{d}\xi$ 是否 $f(x)$ 的原函数？$f(x)$ 的原函数这里就是指处处(而不仅是“几乎处处”)适合关系式 $F'(x)=f(x)$ 的函数 $F(x)$.

(2) 若 $f(x)$ 有原函数 $F(x)$，$f(x)$ 是否可积？

可惜，两个问题的答案均是否定的，为此，我们看两个反例：

(1) $f(x)$ 可积但其不定积分不是其原函数的例子.在区间[0,1]中考虑一个可数的处处稠

密的点集$\{a_n\}$，例如有理点的集合即是一个可数的而且处处稠密的点集. 再看一个收敛的正项级数 $\sum_{k=1}^{\infty} p_k = p < +\infty$. 我们定义一个函数 $f(x)$ 如下：

$$f(x) = \sum_{x \geqslant a_{i_k}} p_{i_k}. \tag{13}$$

即是说，看 $x$ 之左方(包括 $x$)：$[0,x]$包含了哪些 $a_{i_k}$，若 $a_{i_k} \in [0,x]$，则将级数中有相同标号的项 $p_{i_k}$ 列入(13)右方，因为 $\sum_{k=1}^{\infty} p_k = p < \infty$ 是正项级数，故其部分项(尽管是无穷多项)所成的级数(13)仍收敛，从而(13)定义的函数 $f(x)$ 是有意义的. 不但如此，当 $x$ 增加时，适合不等式 $a_{i_k} \leqslant x$ 的 $a_{i_k}$ 更多，因此(13)中还要多加几个项：

$$\begin{aligned} f(x+h) - f(x) &= \sum_{x+h \geqslant a_{i_k}} p_{i_k} - \sum_{x \geqslant a_{i_h}} p_{i_k} \\ &= \sum_{x+h \geqslant a_{i_k} > x} p_{i_k} \geqslant 0. \end{aligned} \tag{14}$$

因此，$f(x)$不减而为单调函数. 从而 $f(x)$可积.

但是 $\int_0^x f(x)\mathrm{d}x$ 不可能是原函数，因为若它是原函数 $F(x)$，当有 $F'(x) = f(x)$. 但单调函数可能有第一类间断点，对于(13)中的 $f(x)$，这些间断点就是$\{a_k\}$，而可以证明 $F(x)$在 $a_k$ 处没有$F'(x)$. 这与 $F(x)$是 $f(x)$的原函数矛盾. $F(x)$不但不是几乎处处可求导，而且在一个稠密集上不可导.

(2) $f(x)$有原函数但本身不可积的例子. 令

$$F(x) = \begin{cases} x^2 \sin\dfrac{1}{x^2}, & x \neq 0, \\ 0, & x = 0. \end{cases} \tag{15}$$

令 $f(x) = F'(x)$，可见 $f(x)$有原函数，但 $f(x)$在 $x=0$ 处是无界的，因此不是可积的.

联系着这两个例子来看微积分的基本定理则应该表述如下：

**定理 7 (微积分的基本定理，达布)** 设

(1) $f(x)$在$(A,B)$上可积；

(2) $f(x)$在$(A,B)$上有原函数 $F(x)$，

则对任一闭区间$[a,b] \subset (A,B)$有

$$F(b) - F(a) = \int_a^b f(x)\mathrm{d}x. \tag{16}$$

证 作任一分划 $P$: $a = x_0 < x_1 < \cdots < x_n = b$，有

$$F(b) - F(a) = \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})] = \sum_{i=1}^{n} F'(\xi_i)(x_i - x_{i-1}),$$
$$x_{i-1} < \xi_i < x_i$$

**注 1** 这个证明是达布给出的，而通常教本上的证明是柯西给出的，他假设了 $f(x)$连续，这样 $f(x)$自然可积．为了证明 $\int_a^x f(t)\mathrm{d}t$ 是 $f(x)$的原函数，他应用了积分中值定理，而这就用到 $f(x)$的连续性．柯西既然限于讨论连续函数，定理 7 中两个条件自然都成立，所以柯西可以说当 $f(x)$为连续时，微积分的基本定理成立．达布因为要考虑可积的 $f(x)$，所以要规定它适合这两个条件．此外，达布则用了微分中值定理即拉格朗日公式，而这只需要 $F'(x)$在$(A,B)$上存在．

**注 2** 这个证明虽然很简洁，却未解决很多问题．因为既已假设 $f(x)$只是可积，则它可能有不连续点．如果我们在一个零测度集上把 $f(x)$改变为 $\tilde{f}(x)$，则 $f(x)$之积分值不变（证明见下一节）：$\int_a^x f(x)\mathrm{d}t = \int_a^x \tilde{f}(t)\mathrm{d}t$．因此如果$\int_a^x f(t)\mathrm{d}t = F(x)$ 是 $f(x)$ 的原函数，则也应有 $F(x) = \int_a^x \tilde{f}(t)\mathrm{d}t$ 而为 $\tilde{f}$的原函数．但是因为 $f(x)\neq\tilde{f}(x)$于一个零测度集上，而在这个集上 $F(x)$同时为 $f(x)$与 $\tilde{f}(x)$的原函数．显然这是不可能的．这样看来在可积函数类中，只能考虑几乎处处意义下原函数．即是说：所谓 $F(x)$是 $f(x)$的原函数．即指几乎处处有 $F'(x)$存在，只是几乎处处有 $F'(x)=f(x)$．现在要问，在这样的理解下，定理 7 是否仍成立．肯定不对，因为有一个著名的反例（我们在第二章就讲过了），这就是著名的赫维赛德函数（只看区间$[-1,1]$）．

$$H(x)=\begin{cases}1, & 0\leqslant x\leqslant 1,\\ 0, & -1\leqslant x<0.\end{cases}$$

很明显 $H'(x)=0$ 几乎处处成立（只差 $x=0$ 一点）．但是当 $x>0$ 时

$$1=H(x)-H(-1)\neq\int_{-1}^x H'(t)\mathrm{d}t=\int_{-1}^x 0\mathrm{d}t=0.$$

当然，读者都会看到，现在的原函数 $H(x)$在 $x=0$ 处不连续．那么要问即令设几乎处处意义下的原函数 $F(x)$是连续的，定理 7 对不对？下一节会看到仍然不对．

还有一点要注意，如果 $f(x)$仅只是可积，$F(x)=\int_a^x f(t)\mathrm{d}t$ 当然是$x$ 的连续函数．我们很容易证明（请读者自己证一下）：若 $x$ 是$f$ 的连续点，则 $F'(x)=f(x)$．证法与通常教本相近，但是不能用积分中值定理．定理 5 告诉我们，$f(x)$可积的充分必要条件是它几乎处处连续．于是 $F'(x)=f(x)$很可能仅仅是几乎处处成立．这样，就可积函数类讨论原函数问题应该只能讨论几乎处处意义下的原函数，而定理 7 却是在处处有 $F'(x)=f(x)$的意义下理解原函数的．因此解决不了问题．

实际上问题很深刻：若 $f(x)$可积，它的连续的原函数是否是$\int_a^x f(t)\mathrm{d}t+C$？这一点大体上是对的，下一节将就 $f(x)$为勒贝格可积（黎曼可积只是它的特例）的情况证明这一点．于是要问，怎样刻画 $f(x)$的连续的原函数$\int$

$b_i,\quad i=1,2,\cdots,n\}$来划分 $\Omega$. 以下,我们也称 $I$ 为一区间.

一维区间$[a,b]$可以分划为有限多个小区间之并,一般 $n$ 维区域则不能分成有限多个闭长方体之并,所以我们先限制考虑 $\Omega$ 为一区间的情况,对于区间 $I=\{x_i,\quad a_i\leqslant x_i\leqslant b_i,\quad i=1,2,\cdots,n\}$,我们自然知道$\mathrm{vol}(I)=\prod\limits_{i=1}^{n}(b_i-a_i)$,而若 $I$ 分划为有限个子区间 $I_j$ 之并,$I=\bigcup\limits_{j=1}^{m}I_j$,而且 $\mathring{I}_j\cap\mathring{I}_k=\varnothing$($\mathring{I}_j$ 指$I_j$ 之内域,即集合$\{a_i<x<b_i,i=1,2,\cdots,n\}$),这个条件指 $I_j$ 与 $I_k$ 只沿长方体的边界相接触. 对于每个定义于 $I$ 上的有界函数$f(x)$,令 $m_j=\inf\limits_{I_j}f(x)$,$M_j=\sup\limits_{I_j}f(x)$,我们也可以定义 $f(x)$在此分划 $P$ 下的上和与下和

$$S_P=\sum_{j=1}^{m}M_jm(I_j),\qquad s_P=\sum_{j=1}^{m}m_j\cdot m(I_j).$$

这里 $m(I_j)=\mathrm{vol}(I_j)$. 易见若分划加细,下和必不减而上和必不增,且对任两个分划 $P'$与$P''$必有

$$s_{P'}\leqslant S_{P''}.$$

于是可得 $f$ 之下积分与上积分:$\underline{A}=\sup\limits_{P}s_P$,$\bar{A}=\inf\limits_{P}S_P$,且 $\underline{A}\leqslant\bar{A}$,若 $\underline{A}=\bar{A}=A$,则称 $f$ 在$I$ 上黎曼可积,这个公共值 $A$ 称为$f$ 在$I$ 上的黎曼积分,记作

$$A=\int_I f(x)\mathrm{d}x\ .$$

关于 $f$ 可积的充分必要条件,本节之引理及定理 1~定理 5 均成立(但推论 6 则不然,因为多元函数之单调性难于定义,定理 7 在 $n$ 维情况下也没有简单的类比),此外我们还需要

**引理 8** 若 $f$ 与$g$ 均为可积,则 $f\cdot g$ 也可积.

**证** 证 $B_f=\{x\in I,f(x)$在 $x$ 点不连续$\}$,$B_g$ 与 $B_{fg}$之意义相似. 容易看到

$$B_{fg}\subset B_f\cup B_g.$$

由假设 $B_f$ 与$B_g$ 均为 0 测度集,可见其并也是 0 测度集,$B_{fg}$也是这样,所以 $fg$ 在$I$ 上可积.

利用这个引理就可以在任一有界集合 $C$ 上定义$f$ 之积分,为此先定义 $C$ 之特征函数

$$\chi_C(x)=\begin{cases}1, & x\in C,\\ 0, & x\notin C.\end{cases}\tag{17}$$

现在设 $C$ 含于某区间$A$ 内:$C\subset A$,若 $f(x)$在 $A$ 上有界,而且$(\chi_C\cdot f)(x)$在 $A$ 上可积,我们就说 $f(x)$在 $C$ 上可积,而且定义

$$\int_C f(x)\mathrm{d}x=\int_A(\chi_C\cdot f)(x)\mathrm{d}x\ .\tag{18}$$

这里有两点值得注意,首先若是选不同的 $A_1$ 与 $A_2$ 均包含 $C$ 则因 $\chi_C|_{A_1}-\chi_C|_{A_2}=0$. 易见 $\int_{A_1}(\chi_C\cdot f)(x)\mathrm{d}x=\int_{A_2}(\chi_C\cdot f)(x)\mathrm{d}x$,所以定义(18)是与 $A$ 的选择无关的. 其次,若$f(x)$本来就定义在 $A$ 上且在$A$ 上可积,则由引理 8,只要 $\chi_C(x)$在 $A$ 上可积,则$(\chi_C\cdot f)(x)$也在 $A$ 上可积,从而可以用(18)式定义$\int_C f(x)\mathrm{d}x$. 所以 $A$ 上一个可积函数$f(x)$在 $A$ 的子集$C$ 上仍可积,只要 $\chi_C(x)$可积. 这样看来 $\chi_C(x)$的积分是特别有意义的.

如果 $C$ 是一个区间：$C=\{x: a_i\leqslant x_i\leqslant b_i, i=1,2,\cdots,n\}$，很明显

$$\int_C \chi_C(x)\mathrm{d}x = \prod_{i=1}^{n}(b_i-a_i) = m(C)$$

是此区间的体积，记作 $m(C)$ 就是前面的 vol $(C)$. 如果 $C$ 是有限多个区间之并：$C=\bigcup_{j=1}^{h} C_j$，而且 $\mathring{C}_j\cap\mathring{C}_k=\varnothing\,(j\neq k)$，可见 $m(C)=\sum_{j=1}^{h} m(C_j)$，它仍然表示 $C$ 的体积. 此式反映了体积的有限可加性. 如果 $C$ 是一般区域，$\chi_C(x)$ 是否可积虽不得而知，但上(下)积分是一定有的，上积分是上和 $\sum_{i=1}^{n} M_i\cdot m(\sigma_i)$ ($\sigma_i$ 是闭区间，$m(\sigma_i)$ 是其体积) 之下确界. 若 $\sigma_i\cap C\neq\varnothing$，$M_i=1$；否则 $M_i=0$. 所以上和为 $\sum_{\sigma_i\cap C\neq\varnothing} m(\sigma_i)$，就是与 $C$ 相交的闭区间的体积和. 这就相当于求圆面积时所用的外切正多边形面积. 上积分 $\overline{I}$ 就是这些“外包多边形”之“面积”的下确界. 同样，下积分是下和 $\sum_{i=1}^{n} m_i\cdot m(\sigma_i)$ 的上确界. 这里 $m_i=1$，若 $\sigma_i\subset C$；否则 $m_i=0$. 所以下和是 $\sum_{\sigma_i\subset C} m(\sigma_i)$ 即含于 $C$ 内的闭区间的体积和，相当于“内接多边形”之“面积”. 因此求 $\underline{I}$ 即求下和的上确界就相当于从“圆内”逐渐增加“内含多边形”之边数而“穷竭”一个圆. $\chi_C(x)$ 可积即 $\underline{I}=\overline{I}$，亦即不论是“内含”还是“外包”结果都相同，都求出了“面积”. 因此，若 $\chi_C(x)$ 可积，就说 $C$ 是有体积的，而体积就是 $m(C)=\int\chi_C(x)\mathrm{d}x$. 但这个积分是体积的一个推广，因为它可以适用到 $C$ 的构造极复杂的情况，所以我们称 $m(C)$ 为 $C$ 的若尔当测度. $\overline{I}$ 与 $\underline{I}$ 也分别称为 $C$ 之外与内若尔当测度，并分别记为 $\overline{m}(C)$ 与 $\underline{m}(C)$. 并不是任何的 $C$ 都有若尔当测度 (或者说 $C$ 为若尔当可测 ($J$ 可测))，因为并不是 $\chi_C(x)$ 在一切情况下均可积. 例如设 $C=\{x: 0\leqslant x\leqslant 1, x$ 为有理数$\}$，$C\subset[0,1]$ 是有界的，若用有限多个区间作 $C$ 的覆盖 $\{(a_i,b_i)\}_{i=1,\cdots,k}$，必有 $\sum_{i=1}^{h}(b_i-a_i)\geqslant 1$，而每一个区间中必含有理数，故若作 $\chi_C(x)$，在任何区间上均有 $M_i=1$，因此上和为 $\sum_{i=1}^{n}(b_i-a_i)\geqslant 1$，其下确界就是 1，反过来也正因为任一区间 $(a_i,b_i)$ 中均有无理数，故必有 $m_i=0$，从而一切下和均为 0，其上确界也就是 0. 所以 $\chi_C(x)$ 不可积，也就是说 $C$ 不是若尔当可测的.

我们要注意，现在的 $\chi_C(x)$ 就是由 (4) 定义的狄利克雷函数 (但限制 $0\leqslant x\leqslant 1$)，所以 $C$ 为若尔当不可测就说明狄利克雷函数 (4) 在 $[0,1]$ 上不是黎曼可积的.

至此我们要问，有界集 $C$ 为若尔当可测的条件是什么？这里我们有

**定理 9**　有界集 $C\subset A$ 为若尔当可测的充分必要条件是：$C$ 的边界点集具有测度 0.

**证**　上面我们已经看到，$C$ 为若尔当可测的充分必要条件是 $\chi_C(x)$ 为黎曼可积. 由定理 3 知，$C$ 为若尔当可测的充分必要条件是 $\chi_C(x)$ 几乎处处连续. 任取

在此开区间中恒为 1，从而 $\chi_C(x)$在 $x_0$ 连续. 若 $x_0$ 为 $C$ 之外点，则必有以 $x_0$ 为中心的开区间全含于 $C$ 之余集内，而在此开区间中 $\chi_C(x)=0$ 在 $x_0$ 仍为连续. 若 $x_0$ 是 $C$ 之界点，则在 $x_0$ 之任一邻域中既有 $C$ 中之点亦有 $C$ 外之点，因此，$O(\chi_C, x_0)=1$，因此{$C$ 之界点}$\subset${$\chi_C(x)$ 连续点}. 反过来的推理也是对的. 因此

$$\{x: x \text{ 为 } C \text{ 之界点}\}=\{x: x \text{ 为 } \chi_C \text{ 之不连续点}\}.$$

再用定理 3 即得.

以上我们看到了如何用有限多个区间之并来"逼近"一个若尔当可测集. 这也决定了若测度的性质. 以下我们称有限多个区间之并为简单集. 若尔当测度中我们最要关切的是其性. 这里我们有

**定理 10**(若尔当测度的有限可加性) 设 $C_1, \cdots, C_m$ 均为若尔当可测集，则 $C=\bigcup_j$ 也是，而且

$$m(C) \leqslant \sum_{j=1}^{m} m(C_j).$$

等号成立的充分必要条件是

$$m(C_i \cap C_j)=0, i \neq j.$$

证 我们只对 $m=2$ 的情况给出证明. 容易证明，当 $C_1, C_2$ 均为简单集(即有限多个闭之并，而任两个相邻的闭区间只交于边界上，下同)时定理成立，而且

$$m(C_1 \cup C_2)=m(C_1)+m(C_2)-m(C_1 \cap C_2).$$

若 $C_1, C_2$ 是若尔当可测的而不一定是简单集，则对任意 $\varepsilon>0$ 必可找到简单集 $Q_1, Q_2$ 分别 $C_1, C_2$ 且

$$m(Q_i) \leqslant m(C_i)+\frac{\varepsilon}{2}, i=1,2.$$

因为 $C_1 \cup C_2 \subset Q_1 \cup Q_2$，对简单集 $Q_1, Q_2$ 应用(21)式有

比较(22)和(23)，注意到 $\underline{m}(C_1\cap C_2)\leqslant\overline{m}(C_1\cap C_2)$，$\underline{m}(C_1\cup C_2)\leqslant\overline{m}(C_1\cup C_2)$，利用 $\varepsilon$ 的任意性，即有

$$\underline{m}(C_1\cup C_2)=\overline{m}(C_1\cup C_2),$$

$$\underline{m}(C_1\cap C_2)=\overline{m}(C_1\cap C_2).$$

即 $C_1\cup C_2$ 与 $C_1\cap C_2$ 均为若尔当可测，而且(21)式成立：

$$m(C_1\cup C_2)=m(C_1)+m(C_2)-m(C_1\cap C_2).$$

定理证毕.

**注**　我们不但证明了 $C_1\cup C_2$ 若尔当可测，而且还证明了 $C_1\cap C_2$ 也若尔当可测. 用类似的方法还可证明 $C_1\setminus C_2$ 也是若尔当可测的，而且

$$m(C_1\setminus C_2)=m(C_1)-m(C_1\cap C_2), \tag{24}$$

(21)与(24)给出了在集合的运算：并、交、差之下，若尔当测度的相应的代数运算结果.

现在我们可以再回来看高维情况下黎曼积分的定义. 上面我们通过采用 $\chi_C(x)$把积分区域 $A$ 化为一个区间，而每一次分划都是把 $A$ 分为有限多个小区间. 现在有了若尔当测度的概念我们可以更直接地从定义在有界区域 $\Omega$ 上的有界函数 $f(x)$着手. 这时唯一需要附加规定的是设 $\Omega$ 为若尔当可测. 我们把 $\Omega$ 分划为有限多若尔当可测集$\Omega_j$，$j=1,2,\cdots,N$，这里我们规定 $\mathring{\Omega}_i\cap\mathring{\Omega}_j=\varnothing\ (i\neq j)$，于是可以作达布下、上和：

$$s_P=\sum_{j=1}^{N}m_j\cdot m(\Omega_j),\qquad m_j=\inf_{\Omega_j}f(x),$$

$$S_P=\sum_{j=1}^{N}M_j\cdot m(\Omega_j),\qquad M_j=\sup_{\Omega_j}f(x).$$

然后把分划加细，证明$\{s_P\}$有上确界，即下积分 $\underline{I}$，$\{S_P\}$有下确界即上积分 $\overline{I}$，唯一需要注意之点是：当分划加细时，$\Omega_j$ 被分划为有限多个“更小”的若尔当可测集之并：$\Omega_j=\bigcup\limits_{k=1}^{L_j}\Omega_{jk}$. 于是由若尔当测度的有限可加性，有

$$m(\Omega_j)=\sum_{k=1}^{L_j}m(\Omega_{jk}).$$

于是

$$m_j\cdot m(\Omega_j)=\sum_{k=1}^{L_j}m_j\cdot m(\Omega_{jk})\leqslant\sum_{k=1}^{L_j}m_{j_k}\cdot m(\Omega_{j_h}).$$

然后由此得知在分划加细时下和不减、而上和不增. 这正是黎曼积分定义中的关键之处. 然后即有可积性之定义：$\underline{I}=\overline{I}$，而前面讲到的各个定理在此均成立. 特别是，若 $f(x)$在 $\Omega$ 上可积，则对任一 $\varepsilon>0$ 必有 $\delta>0$ 存在，使当分划 $P$ 中各个$\Omega_j$ 之直径均小于$\delta$ 时

$$\left|\sum_{j=1}^{N}f(\xi_j)m(\Omega_j)-I\right|<\varepsilon,$$

不论 $\xi$

为了使这个做法有效，必须要求 $\Omega$ 为若尔当可测，以及 $\Omega$ 只能分划为若尔当可测的子集合 $\{\Omega_j\}$. 这样就可以利用若尔当测度的有限可加性. 以后我们会看到黎曼积分的不足之处正好来自于此. 关于测度(面积、体积)，我们直观地会想到把 $\Omega$ 分为可数多个子集之并：$\Omega=\bigcup_{j=1}^{\infty}\Omega_j$，$\mathring{\Omega}_i\cap\mathring{\Omega}_j=\varnothing(i\neq j)$，而且希望

$$m(\Omega)=\sum_{j=1}^{\infty}m(\Omega_j). \tag{25}$$

若一种测度适合(25)式，就说它具有可数可加性. 若尔当测度恰好没有这种可数可加性. 有例为证：考虑[0,1]中的有理数之集 $\Omega$，因为有理数最多有可数多个，所以[0,1]中的有理数可以排列成 $x_1,\cdots,x_N,\cdots$，令 $\Omega_j=\{x_j\}$，很显然 $m_{\Omega_j}=0$. 但 $\Omega$ 则不是若尔当可测的. 因为 $\Omega$ 没有"内容"，故其内测度必为 0：$\underline{m}(\Omega)=0$，但若用有限多个区间覆盖 $\Omega$，$\Omega\subset\bigcup_{i=1}^{N}[a_i,b_i]$，很容易看到 $\sum_{i=1}^{N}(b_i-a_i)\geqslant 1$，所以 $\bar{m}(\Omega)\geqslant 1$(但若许可用可数多个区间覆盖 $\Omega$，则结果反而不同，请读者自己想一下). 因此 $m(\Omega)$不存在，更不必说(25)式可否成立了. 这就说明. 黎曼积分本质地与若尔当测度相联系. 若问黎曼为什么想不到突破这一点？回答只能是：从古希腊到黎曼的两千多年，求面积(例如圆面积)，都是作有限分划(作内接正 $n$ 正边形)再求极限.

在进一步讨论黎曼积分的不足之处前，我们还要再看一下重积分怎样化为逐次积分.

如果考虑 $f(x)$在区间 $A$ 上的积分(若 $f(x)$只定义在 $C\subset A$ 上，则考虑 $\chi_C(x)f(x)$在 $A$ 上的积分)，用小区间来分划 $A$ 是最自然的方法. 以 2 维情况为例，这时 $f(x)$是 $f(x,y)$，$m(\sigma_{ij})=(x_i-x_{i-1})(y_j-y_{j-1})$，而积分和是

$$\sum_{i,j}f(\xi_i,\eta_j)m(\sigma_{ij})=\sum_{i=1}^{L}\Big(\sum_{j=1}^{M}f(\xi_i,\eta_j)(y_j-y_{i-1})\Big)(x_i-x_{i-1}),$$

求极限后，即得到大家熟知的二重积分化为逐次积分的公式

$$\iint_A f(x,y)\mathrm{d}x\mathrm{d}y=\int_a^b\mathrm{d}x\int_c^d f(x,y)\mathrm{d}y. \tag{26}$$

这里的极限过程我们不再重复，我们只提醒一点：当 $f(x,y)$在 $A$ 上连续时这里的论证不会产生困难. 但若 $f(x,y)$只是可积，则情况比较复杂. 因为这时不知道 $f(x,y)$对于固定的 $x$ 是否对 $y$ 可积，但 $\underline{\int}f(x,y)\mathrm{d}y$ 却是一定存在的. 同时也不知道 $\underline{\int}f(x,y)\mathrm{d}y$ 作为 $x$ 的函数是否一定可积. 下面我们把 $A$ 写成 $A=I\times J$，$I=\{x:a\leqslant x\leqslant b\}$，$J=\{y:c\leqslant y\leqslant d\}$. 对 $I$ 和 $J$ 作以下的分划

$$P_I:\ a=x_0<x_1<\cdots<x_n=b,$$

$$P_J:\ c=y_0<y_1<\cdots<y_m=d.$$

这样构成了 $A$ 的一个分划

$$A=\bigcup_{i,j}\sigma_{ij},\ \sigma_{ij}=\{(x,y),x_{i-1}\leqslant x\leqslant x_i,y_{j-1}\leqslant y\leqslant y_j\},$$

$$m(\sigma_{ij})=($$

$$m_{ij}=\inf_{\sigma_{ij}} f(x,y)\leqslant \inf_{y_{j-1}\leqslant y\leqslant y_j} g_x(y)=m_j(g_x(y)),$$

从而

$$\begin{aligned}\sum_j m_{ij}(y_j-y_{j-1})&\leqslant \sum_{j=1}^{m} m_j(g_x(y))(y_j-y_{j-1})\\&\leqslant \underline{\int} g_x(y)\mathrm{d}y.\end{aligned}$$

首先限制 $x_{i-1}\leqslant x\leqslant x_i$，再对右方取下确界，有

$$\sum_j m_{ij}(y_j-y_{j-1})\leqslant \inf_{x_{i-1}\leqslant x\leqslant x_i}\underline{\int} g_{(x)}(y)\mathrm{d}y.$$

双方再乘以 $x_i-x_{i-1}$ 并对 $i$ 求和，即有

$$\sum_{i,j} m_{ij}(x_i-x_{i-1})(y_j-y_{j-1})\leqslant \sum_i \inf_{x_{i-1}\leqslant x\leqslant x_i}\underline{\int} g_x(y)\mathrm{d}y\cdot(x_i-x_{i-1}),$$

左方是 $f(x,y)$ 对分划 $P_1\times P_J$ 的下和，右方是 $\underline{\int} g_x(y)\mathrm{d}y$ 对分划 $P_I$ 的下和，求其上确界即有

$$\underline{\iint} f(x,y)\mathrm{d}x\mathrm{d}y\leqslant \underline{\int}\mathrm{d}x\,\underline{\int} g_x(y)\mathrm{d}y. \tag{27}$$

与此相似，可得

$$\overline{\iint} f(x,y)\mathrm{d}x\mathrm{d}y\geqslant \overline{\int}\mathrm{d}x\,\overline{\int} g_x(y)\mathrm{d}x. \tag{28}$$

合并这两个不等式，注意到我们假设了 $f(x,y)$ 是可积的，即有

$$\begin{aligned}\iint_A f(x,y)\mathrm{d}x\mathrm{d}y &\leqslant \underline{\int}\mathrm{d}x\,\underline{\int} g_x(y)\mathrm{d}y\leqslant \overline{\int}\mathrm{d}x\,\underline{\int} g_x(y)\mathrm{d}y\\&\leqslant \overline{\int}\mathrm{d}x\,\overline{\int} g_x(y)\mathrm{d}y\leqslant \iint_A f(x,y)\mathrm{d}x\mathrm{d}y.\end{aligned}$$

注意，我们在中间插入了一项 $\overline{\int}\mathrm{d}x\,\underline{\int} g_x(y)\mathrm{d}y$. 因为不等式两端的项相同，这就证明了

$$\underline{\int}\mathrm{d}x\,\underline{\int} g_x(y)\mathrm{d}y=\overline{\int}\mathrm{d}x\,\underline{\int} g_x(y)\mathrm{d}y.$$

也就是说，$\underline{\int} g_x(y)\mathrm{d}y$ 作为 $x$ 的函数在 $I$ 上可积，而上式双方均等于 $\int\mathrm{d}x\,\underline{\int} g_x(y)\mathrm{d}y$. 用完全同样的方法也可证明 $\overline{\int} g_x(y)\mathrm{d}y$ 作为 $x$ 的函数在 $I$ 上可积，且 $\underline{\int}\mathrm{d}x\,\overline{\int} g_x(y)\mathrm{d}y=\overline{\int}\mathrm{d}x\,\overline{\int} g_x(y)\mathrm{d}y$.

当然我们可在 $f(x,y)$ 中先固定 $y$ 而得 $x$ 的函数 $f(x,y)=h_y(x)$，它对 $x$ 在 $I$ 上的上、下积分必定也都是 $y$ 在 $J$ 上的可积函数. 总结起来，我们有

**定理 11（富比尼（Fubini）定理）** 设 $f(x,y)$ 在区间 $A=I\times J$ 上可积，则 $f(x,y)$ 作为 $y$（或 $x$）的函数 $g_x(y)$（$h_y(x)$）之上、下积分均在 $J$（或 $I$）上可积，而且 $f(x,y)$ 在 $A$ 上的二重积分可

以化为逐次积分如下：

$$\iint_A f(x,y)\mathrm{d}x\mathrm{d}y = \int_I \mathrm{d}x\underline{\int_J} g_x(y)\mathrm{d}y = \int_I \mathrm{d}x\overline{\int_J} g_x(y)\mathrm{d}y$$

$$= \int_J \mathrm{d}y\underline{\int_I} h_y(x)\mathrm{d}x = \int_J \mathrm{d}y\overline{\int_I} h_y(x)\mathrm{d}x. \tag{29}$$

注 若 $f(x,y)$ 是连续函数，则容易证明 $g_x(y)$，$h_y(x)$ 也都是连续的，从而 $\overline{\int} g_x(y)\mathrm{d}y = \underline{\int} g_x(y)\mathrm{d}y = \int g_x(y)\mathrm{d}y$，$\overline{\int} h_y(x)\mathrm{d}x = \underline{\int} h_y(x)\mathrm{d}x = \int h_y(x)\mathrm{d}x$，这时 (29) 就成了通常的化二重积分为逐次积分的公式

$$\iint_A f(x,y)\mathrm{d}x\mathrm{d}y = \int_I \mathrm{d}x\int_J g_x(y)\mathrm{d}y = \int_J \mathrm{d}y\int_I h_y(x)\mathrm{d}x. \tag{30}$$

但是对于一般可积函数 (30) 甚至可以是没有意义的. 例如取 $I=J=[0,1]$，令 $R(x)$ 为 (5) 式定义的黎曼函数，它在 $x$ 为有理数时不连续. 但这些不连续点构成一个零测度集，所以 $R(x)$ 在 $[0,1]$ 上可积. 读者不妨自己试证 $\int_0^1 R(x)\mathrm{d}x=0$. 令 $\chi(y)$ 是 (4) 式所定义的狄利克雷函数. 上面已经说过，它是不可积的，$\int_0^1 \chi(y)\mathrm{d}y$ 这个记号没有意义. 今令 $f(x,y)=R(x)\chi(y)$，则它只在 $x$ 为有理数时不连续，它的不连续点的集合是可数多个线段之并：

$$\{(x,y): x \text{ 为有理数}, 0\leqslant y\leqslant 1\}.$$

读者自己可以证明它是一个 0 测度集，因此

$$\iint_{I\times J} R(x)\chi(y)\mathrm{d}x\mathrm{d}y$$

是有意义的，而且等于 0. 但 (30) 不可能成立，因为 $\int g_x(y)\mathrm{d}y$ 对有理值 $x$ 没有意义，所以，如果限于黎曼积分，富比尼定理还只能写成定理 11 那样的形状. 读者可能会想到，当我们求一般的区域 $C$ 上的二重积分 $\iint_C f(x,y)\mathrm{d}x\mathrm{d}y$ 时，是取一个较大的区间 $A$，$C\subset A$ 而以 $\iint_A \chi_C(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$ 代替以上积分. 这时例如当 $x=x_0$ 时. 被积函数在 $P_1$，$P_2$ 两点出现不连续点. 不过 $\int_{x=x_0} g_x(y)\mathrm{d}y$ 在 $x=x_0$ 处仍是连续的，所以并不影响 (30) 式之成立 (图 4-1-2). 看来，在研究二重积分时我们时常不自觉地以为区域的构造都比较简单，而当区域比较简单时确实不会出什么问题.

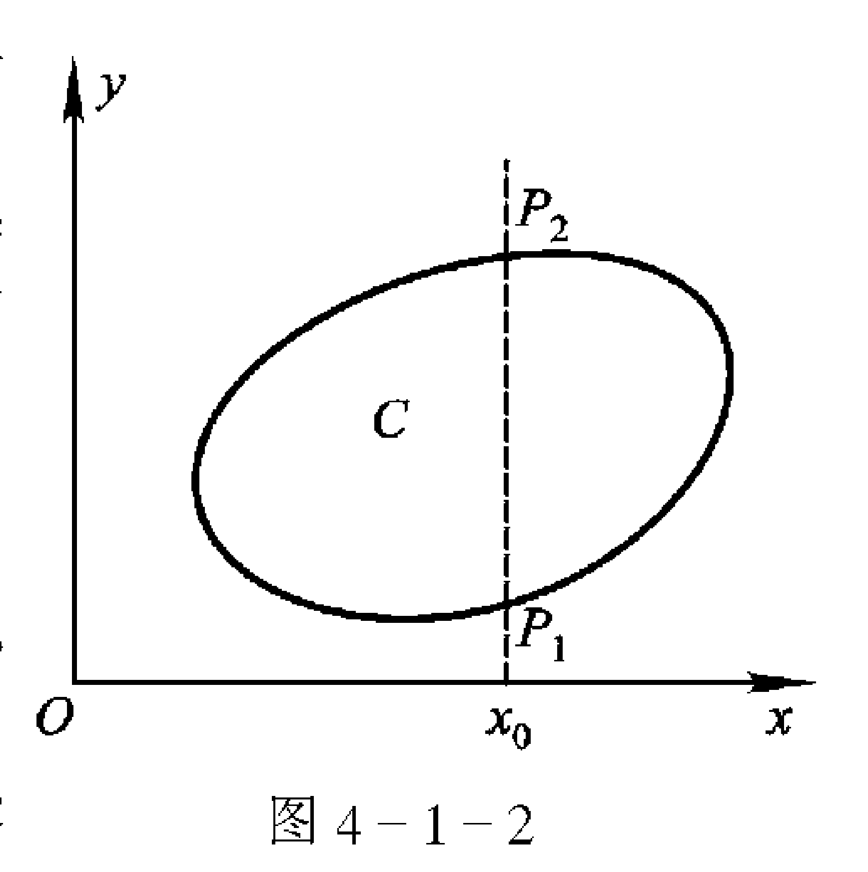


图 4-1-2

**5. 黎曼积分的缺点** 上面我们详尽地讨论了黎曼积分的可积性问题，而且发现其中隐藏着不少常被忽视的问题，例如若尔当测度只是有限可加. 那么，黎曼积分的不足之处究竟何在？如

果改进了有限可加性，这些不足之处是否会得到改进？首先看一个重大的不足之处也可能是最重要的. 以第三章讲的变分问题为例，我们可以找到一个极小化序列$\{u_n\}$使电场能量$E(u_n)$趋向极小：

$$E(u_n)=\iint_\Omega\left[\left(\frac{\partial u_n}{\partial x}\right)^2+\left(\frac{\partial u_n}{\partial y}\right)^2\right]\mathrm{d}x\mathrm{d}y\to\min.$$

黎曼本人就以为这个序列$\{u_n\}$至少有一个子序列确实收敛到某一极限函数$u$，而$E(u)=\min$，于是$u$给出狄利克雷问题之解. 魏尔斯斯拉斯指出这是不对的. 现在$E(u)$是积分号下又出现导数的，如果我们能证明$\{u_n\}$，$\left\{\frac{\partial u_n}{\partial x}\right\}$，$\left\{\frac{\partial u_n}{\partial y}\right\}$均在$\Omega$中一致收敛，则$\{u_n\}$之极限函数$u$确实适合$\frac{\partial u_n}{\partial x}\to\frac{\partial u}{\partial x}$，$\frac{\partial u_n}{\partial y}\to\frac{\partial u}{\partial y}$（在$\Omega$中一致成立），而$u$确实是变分问题之解. 但是，由$E(u_n)\to\min$要证明上述三个序列均一致收敛，确实太难. 实际上它一般是不可能的. 由$E(u_n)\to\min$可以证明，$E(u_n-u_m)\to 0$，而且因为现在有边值条件$u_n|_{\partial\Omega}=g$，还可以证明

$$\iint_\Omega|u_n-u_m|^2\mathrm{d}x\mathrm{d}y\leqslant CE(u_n-u_m)\to 0$$

（这里要用到一个著名的庞加莱不等式，因为超出我们的能力，所以不证了）. 但是即令由$\iint_\Omega|u_n-u_m|^2\mathrm{d}x\mathrm{d}y\to 0$想证明$u_n-u_m\to 0$在$\Omega$内一致成立就已是不可能的事了. 即令我们退一步问，设$\{f_n\}$是连续函数序列，而且$\int|f_n-f_m|\mathrm{d}x\to 0$，可否证明$\{f_n\}$一定有某种意义下的极限$F(x)$使$\int(f_n-F)\mathrm{d}x=0$？或者把问题再说明白一些，可否在积分号下求极限？有例子说明这是不可能的，而这才是黎曼积分的主要缺点. 为了说明这个例子，我们先把序列换成级数，即是设$f_n(x)$是级数$\sum\limits_{n=1}^{\infty}g_n(x)$的$n$项部分和，$g_n(x)$在$[0,1]$上连续. 有例子表明$\sum\limits_{n=1}^{\infty}\int_0^1 g_n(x)\mathrm{d}x\neq\int_0^1\sum\limits_{n=1}^{\infty}g_n(x)\mathrm{d}x$. 实际上$\sum\limits_{n=1}^{\infty}g_n(x)$可能不可积.

但是有一个很有意思的定理：

**定理 12**　若$g_n(x)$在区间$A$上非负可积，而且$\sum\limits_{n=1}^{\infty}g_n(x)$在$A$上收敛，则有

$$\underline{\int}\left[\sum_{n=1}^{\infty}g_n(x)\right]\mathrm{d}x=\sum_{n=1}^{\infty}\int_A g_n(x)\mathrm{d}x.$$

这个定理的证明从略，但是定理本身告诉我们，为了克服黎曼积分不能用于级数的逐项积分，应该扩充积分的定义，使对较宽的一类函数，下积分就是积分. 为了证明此定理又要用到著名的迪尼(Dini)定理如下：

**定理 13（迪尼定理）**　设$g_n(x)$在$A$上非负连续而且$\{g_n(x)\}$下降地趋于0，则$\{g_n(x)\}$必在$A$上一致收敛于0，从而

$$\int_A g_n(x)\mathrm{d}x\to 0.$$

这个定理告诉我们,单调性在积分理论中起了很大的作用,而迄今为止我们没有充分注意到它.但在下一步改进黎曼积分时却不可避免地要注意于此.

现在我们就可以着手来扩充与改进黎曼积分的理论了.这就是引进勒贝格积分.

## §2 勒贝格积分的初步介绍

**1. 勒贝格积分理论出现的背景** 上一节中我们详尽地讨论了黎曼积分的可积性问题,指出了它的不足之处,而且种种迹象表明,其局限性来自(或者说反映在)若尔当测度只有有限可加性.我们还指出了,迄今我们都没有充分注意单调性在积分理论中起的作用.自20世纪初,法国数学家勒贝格(H. Lebesgue)从研究测度着手提出了一种新的积分理论,即所谓勒贝格积分,有效地克服了黎曼积分的不足之处①.

与数学的发展同时,物理学也在发展,二者表面上互相独立,但实际上有深刻的联系.这里谈到的物理学的发展是指统计的思想进入了物理学.直到19世纪末年,原子究竟是不是实际存在,在物理学家中还是一个问题.当时即已有一批物理学家认为气体乃是许多粒子的总体.这些粒子就是原子(那时对原子和分子的区别也还不清楚).它们随机地运动,互相碰撞,改变速度.气体的许多宏观的性质都可以用原子的运动来解释,例如压力、粘性等等.这就是气体的分子运动理论.其实更早提出这个观点的还有18世纪的丹尼尔•伯努利(Daniel Bernoull.居住在瑞士的巴塞尔的伯努利家族是一个数学世家.从这个家族出来了一大批杰出的数学家,丹尼尔已经是第三代了).不过这批物理学家在麦克斯韦(James Clerk Maxwell)以前,都认为这些分子具有相同的速度.到麦克斯韦才知道这是不对的.于是形成了一个观点,即我们不能问某一个气体分子在某一时刻速度究竟是多少,而只能问:一个分子速度在 $u$ 到 $u+\mathrm{d}u$ 间的概率(即这类分子占分子总数的百分比)是多少.麦克斯韦从一些类比、或数学上的考虑指出,这个概率是 $f(u)\mathrm{d}u$,而 $f(u)=\sqrt{m/2\pi kT}\exp(-mu^2/2kT)$,称为概率密度.所以,麦克斯韦给出了分子速度的一种分布,这种分布现在就称为麦克斯韦分布或正态分布、高斯分布.它在随机现象的研究中极为重要.其后,奥地利物理学家玻耳兹曼(Ludwig Boltzmann,1844—1906)和美国物理学家吉布斯(Josiah Willard Gibbs,1839—1903)都作出了重大贡献.使得人们理解了,随机性是来自物理世界的本质.因此,必须从统计随机的观点来研究物理现象.例如,当20世纪初,控制论的创始人维纳(Norbert Wiener,1894—1964)着手研究吉布斯的理论时,就发现了勒贝格积分理论正是研究随机现象的最好的数学工具.他发现,概率从本质上说就是一种测度,而勒贝格积分也正是始于对测度的研究.

于是我们看到,数学的发展和物理学的发展又互相推动.勒贝格积分和测度论成了数学物理以及许多数学分支的基础,它们是数学中一个广大的新领域.本书中不可能对勒贝格积分理论作详细的介绍.下面只限于从它克服黎曼积分理论的不足之处这个角度出发,对其要点作简单介绍.

黎曼积分主要适用范围是比较规则的函数.上节即已指出 $f(x)$ 在有限区间 $A$ 上黎曼可积的充分必要条件是 $f(x)$ 几乎处处连续.直观地看,这是容易理解的:若将区间分划成为小区间,则一定要求 $f(x)$ 在其上变动剧烈的小区间,总体说来不太大.这样,用 $m_i, M_i$ 来代替 $f(\xi_i)$ 时,

① 关于勒贝格积分的历史,请参看 J.-P. Kahane 的《Lebesgue 积分的产生及其影响》,数学进展,2002,31(2),97~106

总体说来影响也不太大.但是若 $f(x)$ 在 $A$ 上剧烈振动(图 4－2－1),则在任意小区间上用 $\xi_i$ 来代替 $\xi$ 都会造成积分和很大的变化,使得用积分和之"极限"来作为定积分之定义遭到很大的困难.于是勒贝格提出,不妨在 $y=f(x)$ 的值域上作分划,即把 $[m,M]$ 分成 $m=\eta_0<\eta_1<\cdots<\eta_n=M$ 然后看使得 $f(x)$ 之值落在 $(\eta_{i-1},\eta_i)$ 中的 $x$ 之集合

$$E_i=\{x:\eta_{i-1}<f(x)\leqslant\eta_i\} \tag{1}$$

之"长度" $m(E_i)$.这样就可以得到 $y=f(x)$ 在 $x\in[0,1]$ 上的一段与 $x$ 轴,$x=0$,$x=1$ 所包围的面积(即曲边梯形面积.不过我们不说曲边梯形,因为它的"边"太不规则了,不是一个曲字能描述的)近似值为

$$I\approx\sum_{i=1}^{n}\eta_i m(E_i), \tag{2}$$

然后再求极限.这样做当然解决了 $f(x)$ 在任一个小区间中剧烈振荡的问题,因为现在每一个 $E_i$ 中 $f(x)$ 的变化限于 $[\eta_{i-1},\eta_i]$ 中,但是当 $f(x)$ 很不规则时,$E_i$ 的构造也变复杂了.如图 4－2－1,$E_2$ 由六个小区间组成,于是我们要处理的函数之复杂已非通常图形所能表示.从上一章§7 讲到佩韩对布朗运动的轨道之评论可以想到这一点.至少我们应该考虑 $E_i$ 由可数多个互相分离的子集拼成的情况:$E_i=\bigcup_{k=1}^{\infty}E_i^{(k)}$,即令每个 $E_i^{(k)}$ 简单,以至于我们可以定义 $m(E_i^{(k)})$,那么能否规定

$$m(E_i)=\sum_{k=1}^{\infty}m(E_i^{(k)})?$$

至少,若用若尔当测度为 $m$ 是不行的.因为若尔当测度只有有限可加性.而上式表明,我们所需的测度,必须有可数可加性.所以实现勒贝格的想法的第一步是要推广若尔当测度为更一般的,具有可数可加性的测度.

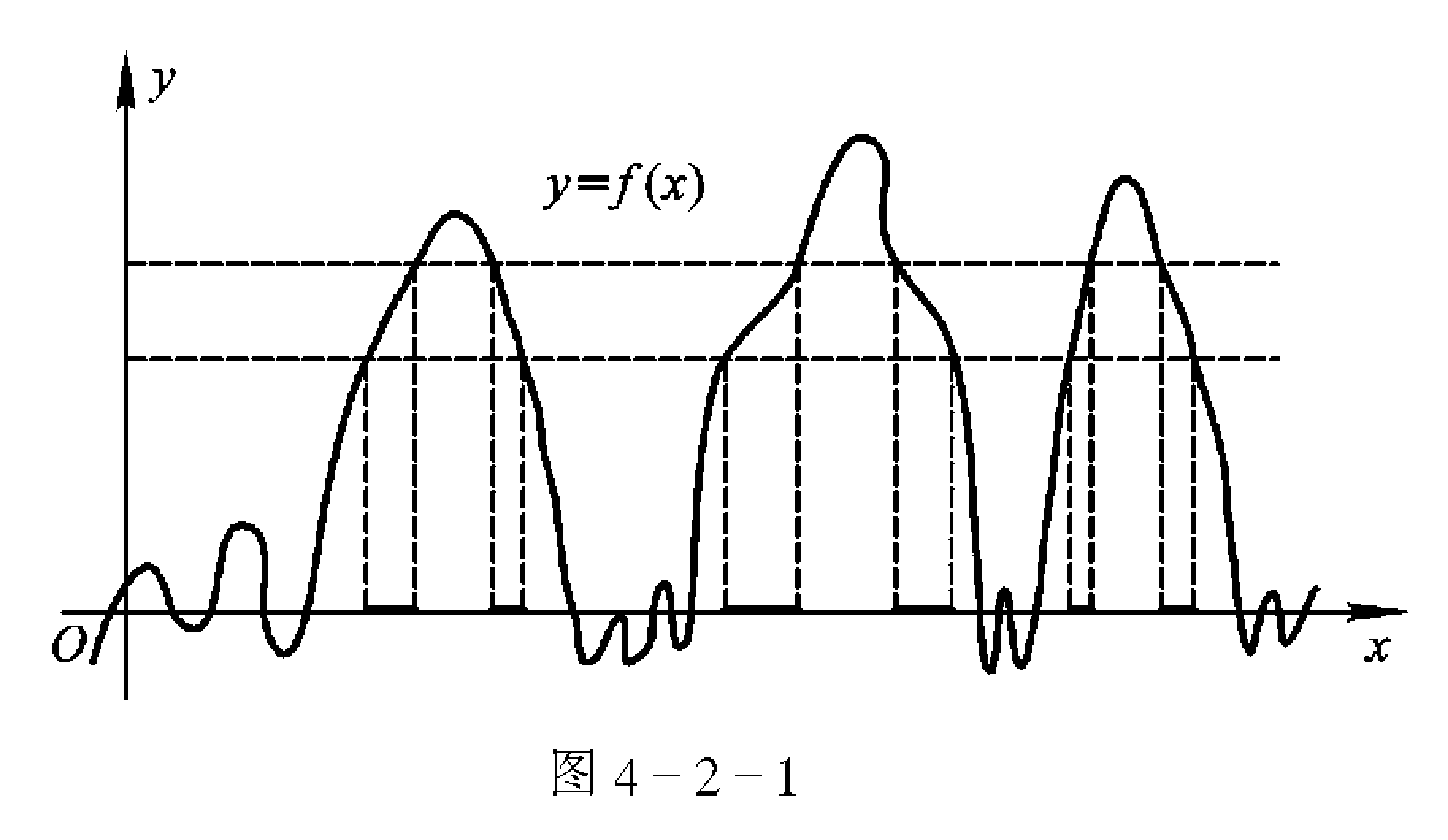



图 4－2－1

**2. 集合的勒贝格测度**

$$m(E)=\sum_{i=1}^{\infty}(b_i-a_i).$$

至此我们已看到，在最简单的情况下，$L$ 测度就是长度(在高维情况下就是体积)，而且已经预置了可数可加性于其中.现在可以来看 $E\subset[a,b]$为一个相当一般的集的情况，这时就要把用外切与内接多边形来求圆面积的思想加以发挥了：作包含 $E$ 的开集$F$，而 $F$ 是可个区间 $I_1,\cdots,I_k,\cdots$之并；$F=\bigcup_{k=1}^{\infty}I_k$，而且适合 $\sum_{k=1}^{\infty}m(I_k)<\infty$，令 $m_e(E)=\inf\sum_{k=1}^{\infty}m(I_i$
称之为 $E$ 的外测度.很明显，因为$[a,b]\supset E$，而 $m([a,b])=b-a$，所以

$$0\leqslant m_e(E)\leqslant b-a.$$

$E$ 在$[a,b]$中的余集，记为 $E^c$，我们也可以求 $E^c$ 之外测度$m_e(E^c)\leqslant b-a$，而定义 $E$ 之内为

$$m_i(E)=(b-a)-m_e(E^c).$$

这里我们注意到，$m_i(E)\leqslant m_e(E)$.事实上，由外测度之定义，对任意 $\varepsilon>0$，必可找到如上述与 $F'$，使

$$m(F)\leqslant m_e(E)+\varepsilon,$$
$$m(F')\leqslant m_e(E^c)+\varepsilon.$$

$[a,b]$中任一点均含于 $F$ 或$F'$之某一开区间内，而由这些开区间中可以选出有限个覆盖$[a$
记这些区间为 $I_1,\cdots,I_N$，则

$$\sum_{k=1}^{n}m(I_k)\geqslant b-a,$$

$$\sum_{k=1}^{n}m(I_k)\leqslant m(F)+m(F'),$$

故

$$b-a\leqslant m_e(E)+m_e(E^c)+2\varepsilon.$$

$$m(E)+m(E^c)=b-a. \tag{7}$$

于是我们得到如下的

**定理 1**　令 $E$ 是含于区间 $[a,b]$ 中的集合，则 $E$ 与 $E^c=[a,b]\setminus E$ 必同时可测或同时不可测且

$$m(E)+m(E^c)=b-a.$$

在进一步讨论集的运算与其可测性的关系之前，我们可先提几个明显的结论.

**命题 1**　闭区间 $[\alpha,\beta]\subset[a,b]$ 与半闭区间 $(\alpha,\beta]$，$[\alpha,\beta)$，开区间 $(\alpha,\beta)$ 同样可测，且测度同为 $\beta-\alpha$.

**命题 2**　$[a,b]$ 内之 0 测度集均为可测集，而且其测度为 0.

**证**　设 $E$ 为 0 测度集. 由定义，对任意 $\varepsilon>0$ 均可找到可数多个闭区间 $U_1,\cdots,U_k,\cdots$，使 $E\subset\bigcup_{k=1}^{\infty}U_k$，而且 $\sum_{k=1}^{\infty}\mathrm{vol}(U_k)<\frac{\varepsilon}{2}$. 例如令 $U_k=[\alpha_k-\delta_k,\alpha_k+\delta_k]$，长度为 $2\delta_k$，作 $I_k=(\alpha_k-2\delta_k,\alpha_k+2\delta_k)$，则 $m(I_k)=4\delta_k=2\mathrm{vol}(U_k)$ 而 $\sum_{k=1}^{\infty}m(I_k)=2\sum_{k=1}^{\infty}\mathrm{vol}(U_k)<\varepsilon$. 然而，$E\subset\bigcup_{k=1}^{\infty}U_k\subset\bigcup_{k=1}^{\infty}I_k$，所以由外测度之定义 $m_e(E)<\varepsilon$，由 $\varepsilon$ 之任意性知 $m_e(E)=0$，但任何集均有非负的内测度 $m_i(E)\geqslant 0$，由 $m_i(E)\leqslant m_e(E)$ 知，$0=m_i(E)=m_e(E)$. 证毕.

现在我们来讨论集运算下测度的运算. 首先我们给出具有基本重要性的

**定理 2**（测度的可数可加性）　设 $E_1,\cdots,E_k,\cdots$ 为有限区间 $[a,b]$ 中的可测集. 则

(i)　$E=\bigcup_{k=1}^{\infty}E_k$ 也是可测集，而且

$$m(E)\leqslant\sum_{k=1}^{\infty}m(E_k). \tag{8}$$

(ii)　若再规定 $E_i\cap E_j=\varnothing$，$i\neq j$，则有

$$m(E)=\sum_{k=1}^{\infty}m(E_k). \tag{9}$$

这个定理将与以下定理同时证明.

**定理 3**　设 $E_1,\cdots,E_k,\cdots$ 为有限区间 $[a,b]$ 中的可测集，则

$$E=\bigcap_{k=1}^{\infty}E_k$$

也是可测的.

这两个定理的证明较长，分成几步进行.

(A) 先证若开集 $O=\bigcup_{k=1}^{\infty}O_k$，$O_k$ 均为开集，则

$$m(O)\leqslant\sum_{k=1}^{\infty}m(O_k). \tag{10}$$

开集为最多可数多个开区间之并，若它们互相分离，由 (3) 知它为可测. 又，我们可设 (10) 之右方是收敛的，否则 (10) 式是不足道的，设 $O_k=\bigcup_{m=1}^{\infty}(a_{m,k},b_{m,k})$，$O$ 则由开区间 $(A_l,B_l)$ 并成：

$O=\bigcup\limits_{l=1}^{\infty}(A_l,B_l)$，令 $\varepsilon<\frac{1}{2}(B_l-A_l)$，则 $[A_l+\varepsilon,B_l-\varepsilon]$ 之每一点为某个 $(a_{m,k},b_{m,k})$ 之内点，所以可以在 $(a_{m,k},b_{m,k})$ 中选出有限多个(记其集为 $S$)来覆盖 $[A_l+\varepsilon,B_l-\varepsilon]$ 从而

$$B_l-A_l-2\varepsilon\leqslant\sum(b_{m,k}-a_{m,k}).$$

由 $\varepsilon$ 之任意性，知上式左方的 $\varepsilon$ 可以略去. 对 $l$ 求和即有：

$$\begin{aligned}m(O)&=\sum_l(B_l-A_l)\leqslant\sum_{m,k}(b_{m,k}-a_{m,k})\\&=\sum_k\sum_m(b_{m,k}-a_{m,k})=\sum_k m(O_k).\end{aligned}$$

故(10)得证.

若小区间 $(a_{m,h},b_{m,h})$ 互相均不相交，则 $(A_l,B_l)$ 只是 $(a_{m,k},b_{m,k})$ 以另一次序排列，从而(10)中有等号成立.

(B)若开集 $O,O'$ 覆盖 $[a,b]$ 则

$$m(O\cap O')\leqslant m(O)+m(O')-(b-a). \tag{11}$$

可以从构成 $O$ 的开区间中选用有限多个(记其集为 $Q$)，从 $O'$ 中选取 $Q'$ 也由有限多个区间构成，使 $Q\cup Q'\supset[a,b]$，而且

$$O=Q\cup R,\qquad O'=Q'\cup R',m(R)<\varepsilon,m(R')<\varepsilon.$$

但是

$$O\cap O'\subset(Q\cap Q')\cup R\cup R',$$

从而

$$m(O\cap O')\leqslant m(Q\cap Q')+m(R)+m(R')\leqslant m(Q\cap Q')+2\varepsilon.$$

但因 $Q$ 与 $Q'$ 中都只有有限个区间，从而

$$m(Q\cap Q')\leqslant m(Q)+m(Q')-(b-a)\leqslant m(O)+m(O')-(b-a),$$

代入上式，由 $\varepsilon$ 之任意性即得(11)式.

(C) 对外测度证明(10)式. 因为 $E=\bigcup\limits_{k=1}^{\infty}E_k$，对每个 $E_k$ 可以找到开集 $O_k$ 包含 $E_k$ 而 $m(O_k)<m_e(E_k)+\varepsilon/2^k$，记 $O=\bigcup\limits_{k=1}^{\infty}O_k$，则 $O$ 为开集，利用(10)有

$$m(O)\leqslant\sum_k m(O_k)\leqslant\sum_k m_e(E_k)+\varepsilon.$$

但 $O\supset E$，故 $m(O)\geqslant m_e(E)$，代入上式，令 $\varepsilon\to0$ 即得.

(D) 当 $E_k$ 互不相交时，定理 2 的证明. 先设 $E_1,\cdots,E_k,\cdots$ 只有有限多个. 对于 $E=E_1\cup E_2$，我们已知

$$m_e(E)\leqslant m_e(E_1)+m_e(E_2)=m(E_1)+m(E_2),$$

下面只需证明 $m_i(E)\geqslant m(E_1)+m(E_2)$，亦即

$$m_e(E^c)\leqslant m(E_1^c)+m(E_2^c)-(b-a) \tag{12}$$

即可. 为此分别作包含 $E_1^c$ 与 $E_2^c$ 的开集 $O_1$ 与 $O_2$ 并使

$$m(O_1)\leqslant m(E_1^c)+\varepsilon,m(O_2)\leqslant m(E_2^c$$

$$m(O_1\cap O_2)\leqslant m(O_1)+m(O_2)-(b-a).$$

但 $O_1\cap O_2\supset E_1^c\cap E_2^c=(E_1\cup E_2)^c=E^c$，所以

$$\begin{aligned}m_e(E^c)&\leqslant m(O_1)+m(O_2)-(b-a)\\&\leqslant m_e(E_1^c)+m_e(E_2^c)-(b-a)+2\varepsilon.\end{aligned}$$

令 $\varepsilon\to 0$ 即得(12). 因此 $E=E_1\cup E_2$ 为可测且

$$m(E)=m(E_1)+m(E_2).$$

由此，对有限多个 $E_1,\cdots,E_n$ 也有 $E=\bigcup\limits_{k=1}^{n}E_k$ 为可测，且

$$m\left(\bigcup_{k=1}^{n}E_k\right)=\sum_{k=1}^{n}m(E_k).$$

余下就要看 $n\to\infty$ 的极限情况. 记 $S_n=\bigcup\limits_{k=1}^{n}E_k$，则由上所述已知 $S_n$ 是可测的，且 $\sum\limits_{k=1}^{n}m(E_k)=m\left(\bigcup\limits_{k=1}^{n}E_k\right)=m(S_n)\leqslant b-a$，从而 $\sum\limits_{k=1}^{\infty}m(E_k)$ 收敛，其和 $\leqslant b-a$.

由关于外测度的证明(C)可知

$$m_e(E)\leqslant\sum_{k=1}^{\infty}m_e(E_k)=\sum_{k=1}^{\infty}m(E_k)\leqslant b-a, \tag{13}$$

所以余下的只需证明

$$m_i(E)\geqslant\sum_{k=1}^{\infty}m(E_k) \tag{14}$$

即可. 证如下：

$S_n\subset E$，故 $E^c\subset S_n^c$，但 $S_n$ 作为有限多个可测集 $S_1,\cdots,S_n$ 之并，前已证明为可测的，故其余集亦然. 即

$$m_e(E^c)\leqslant m_e(S_n^c)=m(S_n^c)=(b-a)-\sum_{k=1}^{n}m(E_k).$$

令 $n\to\infty$，由 $\sum\limits_{k=1}^{\infty}m(E_k)$ 收敛即得(14)式.

(E) 有限多个 $E_k$ 时定理 3 的证明. 因为 $E_1,\cdots,E_n$ 可测等价于 $E_1^c,\cdots,E_n^c$ 可测. 但上面的证明已告诉我们，$E_1^c\cup\cdots\cup E_n^c$ 为可测(各项不相交时并集之测度等于各项测度之和这一点虽不成立，我们也未用到)，所以

$$(E_1^c\cup\cdots\cup E_n^c)^c=E_1\cap\cdots\cap E_n$$

也可测. 至于无限多个 $E_k$ 之交的情况将在后面给出. 目前我们先给出一个

**推论 4** 若 $E_1$ 与 $E_2$ 均为可测，则 $E_1\setminus E_2$ 与对称差 $E_1\triangle E_2$ 也可测. 若 $E_2\subset E_1$，则 $m(E_1\setminus E_2)=m(E_1)-m(E_2)$.

证 $E_1\setminus E_2=E_1\cap E_2^c$，故当 $E_1,E_2$ 为可测时 $E_1\setminus E_2$ 也可测. 对称差 $E_1\triangle E_2=(E_1\setminus E_2)\cup(E_2\setminus E_1)$ 是两个不相交的可测集之并，所以是可测的. 最后

$$E_1=(E_1\cap E_2)\cup(E_1\cap E_2^c)=(E_1\cap E_2)\cup(E_1\setminus E_2).$$

当 $E_2\subset E_1$ 时，$E_1\cap E_2=E_2$，再用(D)即得推论.

**推论 5** 空集$\varnothing$可测而且其测度为 0.

**证** 任取可测集 $E$，$\varnothing = E \setminus E$. 由推论 4 即得.

(F) 证明的终结. 令

$$E_1' = E_1, E_2' = E_2 \cap E_1^c, \cdots$$
$$E_n' = E_n \cap (E_1' \cup E_2' \cup \cdots \cup E_{n-1}')^c, \cdots.$$

显然

$$E = E_1' \cup E_2' \cup \cdots \cup E_n' \cup \cdots,$$

而且 $E_i' \cap E_j' = \varnothing\ (i \neq j)$. 所以，应用前面的证明(E)可知 $E$ 是可测的. 关于测度的不等式，则可由关于外测度的证明(C)得出. 因为这时外测度就是测度. 定理 2 至此证毕.

定理 3 则可以由关系式

$$\left(\bigcap_{k=1}^{\infty} E_k\right)^c = \bigcup_{k=1}^{\infty} E_k^c$$

以及定理 2 直接得出.

上面的证明中我们用到"集合的极限过程"，这在有关测度的问题中是很有用的. 为此我们给出

**定义 2** 若有一串集 $E_1, \cdots, E_n, \cdots$，而且 $E_1 \subset E_2 \subset \cdots \subset E_n \subset \cdots$ (或 $E_1 \supset E_2 \supset \cdots \supset E_n \supset \cdots$)，则 $E = \bigcup_{k=1}^{\infty} E_k$ (或 $E = \bigcap_{k=1}^{\infty} E_k$) 称为其外极限(内极限).

由上面的定理易见可测集序列的内外极限均为可测集，且

$$\lim_{n \to \infty} m(E_n) = m(E). \tag{15}$$

在数学中有种种不同的测度. 例如在第二章 §5 讲到广义函数时，就指出 $\delta$ 函数也是一种测度. 在考虑集合 $X$ 的子集时，有一些子集是有测度的——可测集，有一些则没有，我们把 $X$ 之可测的子集作为一类，则可测集的这一类一定要有以下的性质：

(i) $X$ 本身是可测的；

(ii) 若 $E$ 可测，则 $E^c = X \setminus E$ 也可测；

(iii) 若 $E_1, \cdots, E_n, \cdots$ 可测，则 $\bigcup_{n=1}^{\infty} E_n$ 也可测.

我们记这一类子集之集合为 $\mathscr{M}$，则所谓测度就是 $\mathscr{M}$ 上的一个可数可加集合函数 $\mu$，即对 $E \in \mathscr{M}$ 有 $\mu(E) < +\infty$ (称为 $E$ 之测度). 而且若 $E_1, \cdots, E_n, \cdots \in \mathscr{M}$ 而且互不相交，则

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n). \tag{16}$$

所以，勒贝格测度是适合以上条件的测度. 但若尔当测度则不然，因为它没有可数可加性而只有有限可加性(有时也称它为有限可加测度).

**3. 可测函数** 有了勒贝格测度以后就可以进一步定义勒贝格积分了. 第一节中我们一再强调黎曼积分本质上是连续函数(关于若尔当测度)的积分. 那么现在连续函数应该代以什么样的函数类呢？连续函数是这样一个映射：它使开集的原像仍为开集. 于是一个很自然的选择应该是考虑使某个开集的原像为可测集的函数(即可测函数)类. 这正是勒贝格的选择. 下面给出准确的定义，这里我们限于 $f(x)$ 为有界的情况.

**定义 3**　若 $f(x)$ 是定义在 $[a,b]$ 上的函数，而且对任一实数 $c$，集合

$$E(f>c)=\{x;\ x\in[a,b],f(x)>c\} \tag{17}$$

为勒贝格可测集，则称 $f(x)$ 为勒贝格可测函数.

**注 1**　定义中(17)式也可以改为 $E(f\geqslant c)$，$E(f<c)$ 或 $E(f\leqslant c)$，因为例如 $E(f\leqslant c)=[E(f>c)]^c$，故与 $E(f>c)$ 同为可测或为不可测. 又若 $E(f>c)$ 对一切 $c$ 均为可测，则 $E\left(f\leqslant c-\dfrac{1}{n}\right)$ 对一切 $c$ 与正整数 $n$ 均为可测. 由于可数多个可测集之并仍为可测，故 $\bigcup\limits_{n=1}^{\infty}E\left(f\leqslant c-\dfrac{1}{n}\right)=E(f<c)$ 为可测. 其余集 $E(f\geqslant c)$ 当然也可测.

**注 2**　若 $f$ 是可测的，则对任意 $a<b$，$E(a<f<b)$，$E(a<f\leqslant b)$，$E(f=a)$ 均为可测. 证明很简单，例如 $E(a<f<b)=E(f>a)\cap E(f<b)=E(f>a)\cap[E(f\geqslant b)]^c$，$E(f=a)=E(f\geqslant a)\setminus E(f>a)$，当然都是可测的.

可测函数的性质有一些很简单，例如说 $f$ 为可测函数则对任意实数 $h$，$hf$ 也是可测的，这一些我们都不加证明. 但有一些涉及不等式应该注意，例如

**命题 3**　若 $f$ 与 $g$ 均可测则 $F(f>g)=\{x:x\in[a,b],f(x)>g(x)\}$，也是可测集.

**证**　若在某一点 $x$，$f(x)>g(x)$，则必可找到一个有理数 $r$ 使 $f(x)>r>g(x)$. 反之亦然，所以 $E(f>g)=\bigcup\limits_{r=\text{有理数}}E(f>r>g)=\bigcup\limits_{r=\text{有理数}}E(f>r)\cap E(g<r)$ 是可测的.

**命题 4**　若 $f$ 与 $g$ 均为可测，则 $f\pm g$ 与 $fg$ 也是可测的.

**证**　因为 $g$ 可测，则 $-g$ 也可测，常数 $c$ 与它的和当然也是可测的，但

$$E(f\pm g>c)=E(f>c\mp g),$$

故由命题 3，$f\pm g$ 是可测的.

又若 $f$ 可测，$f^2$ 必可测. 实际上

$$E(f^2>c)=[a,b]，若\ c\leqslant 0.$$

$$E(f^2>c)=E(f>\sqrt{c})\cup E(f<-\sqrt{c})若\ c>0.$$

所以 $f^2$ 可测. 利用明显的关系式

$$fg=\frac{1}{4}[(f+g)^2-(f-g)^2]$$

以及前证即知 $fg$ 可测.

一个十分重要的性质是关于极限的. 设有一个可测函数序列{

于是由可测函数的定义以及定理 2,3 知 $E_{m,n}$ 与 $E_m$ 均为可测集. 对于 $x\in E_m$,对任意 $n$ 都可找到一个 $k$ 使 $f_{n+k}(x)>c+\dfrac{1}{m}$,即有充分大的 $\nu$ 使 $f_\nu(x)>c+\dfrac{1}{m}$. 所以在 $E_m$ 上,$\overline{\lim\limits_{n\to\infty}}f_n(x)\geqslant c+\dfrac{1}{m}>c$. 令 $E=E_1\cup E_2\cup\cdots\cup E_n$,$E$ 当然是可测集,今证 $E=\{x:\overline{\lim\limits_{n\to\infty}}f_n(x)>c\}$. 实际上上面已证明,在 $E$ 上,$\overline{\lim\limits_{n\to\infty}}f_n(x)>c$,反之,若有一点 $x$ 使 $\overline{\lim\limits_{n\to\infty}}f_n(x)>c$,则必对充分大整数 $m$,都可以找到充分大的 $\nu$ 使 $f_\nu(x)>c+\dfrac{1}{m}$,所以任取一个 $n$ 都可以找到 $k$ 使 $f_{n+k}(x)>c+\dfrac{1}{m}$. 因此这个 $x\in E_{m,n}$,$n=1,2,\cdots$,从而 $x\in E_m\subset E$. 即是说,若 $x$ 使 $\overline{\lim\limits_{n\to\infty}}f_n(x)>c$,必有 $x\in E$. 所以 $E=E(\overline{\lim\limits_{n\to\infty}}f_n(x)>c)$. 但上面已说了 $E$ 是可测集,所以 $\overline{\lim\limits_{n\to\infty}}f_n(x)$ 也可测. $\underline{\lim\limits_{n\to\infty}}f_n(x)$ 与 $f(x)$ 的可测性证明类似.

**定理 7** 连续函数 $f(x)$ 必为可测.

**证** $E(f>c)$ 即值域空间中之开集 $\{f>c\}$ 在定义域 $x$ 的空间 $[a,b]$ 上之原像. 因为 $f$ 连续,故这个原像也是开集. 开集自然是可测的.

这个定理的重要性在于,我们实际上遇到的函数多是连续函数序列的极限,甚至如狄利克雷函数 $\chi(x)$ 也可写成连续函数的序列如下:

$$\chi(x)=\lim_{m\to\infty}\lim_{n\to\infty}[\cos(m!\,\pi x)]^{2n} \tag{18}$$

($m,n$ 取正整数值). 因为若 $x$ 是有理数 $\dfrac{p}{q}$,则当 $m>q$ 时,$\cos(m!\pi x)=\cos\left(p\cdot\dfrac{m!}{q}\cdot\pi\right)=\pm1$,因此当 $m$ 充分大时,$[\cos(m!\pi x)]^{2n}=1$,而上式之极限也是 1. 若 $x$ 是无理数,不论怎样取正整数 $m$,$m!\pi x$ 均非 $\pi$ 之整数倍,因而 $(\cos m!\pi x)^2=\theta<1$,固定了 $x$ 和 $m$ 以后,$[\cos(m!\pi x)]^{2n}=\theta^n$,而 $\lim\limits_{n\to\infty}[\cos(m!\pi x)]^{2n}=0$,再对 $m$ 取极限为 0. 因为 $[\cos m!\pi x]^{2n}$ 是连续函数,对 $m,n$ 求逐次极限(先对 $n$ 求极限),得出的极限函数即狄利克雷函数自然是可测的. 当然,证明 $\chi(x)$ 可测最好的方法仍然是直接用定义去证明. 上面给出(18)式只是录以备考.

关于可测函数序列的收敛性有许多重要的定理. 定理 6 讲的是它的几乎处处收敛性. 还有许多其它的收敛性,互相之间都有密切的关系,其中最有用的结果应该是下述的叶果洛夫(D. F. Egorov)定理.

**定理 8(叶果洛夫定理)** 设可测集 $E$ 上的可测函数序列 $\{f_n(x)\}$ 几乎处处收敛于几乎处处取有限值的函数

$$R_n(\sigma)=\bigcup_{k=n}^{\infty}E_k(\sigma),$$

$$M=\bigcap_{n=1}^{\infty}R_n(\sigma).$$

很明显 $R_1(\sigma)\supset\cdots\supset R_n(\sigma)\supset\cdots$，所以 $M$ 是 $\{R_n(\sigma)\}$ 的内极限.而有

$$\lim_{n\to\infty}m(R_n(\sigma))=m(M). \tag{20}$$

可以证明 $M\subset Q$，这是因为，若 $x_0\notin Q$，即 $\lim\limits_{k\to\infty}f_k(x_0)=f(x)$，因此，对 $\sigma$ 必有 $N$ 存在，当 $k\geqslant N$ 时

$$|f_k(x_0)-f(x_0)|<\sigma.$$

换言之 $x_0\notin E_k$，只要 $k\geqslant N$. 也就是说 $x_0\notin R_N(\sigma)$ 从而 $x_0\notin M$. 因为已设 $Q$ 为零测度集，故其子集 $M$ 也是零测度集，而由(20)式有

$$\lim_{n\to\infty}m(R_n(\sigma))=0.$$

但 $E_n(\sigma)$ 是 $R_n(\sigma)$ 之子集，故对任意 $\sigma$

$$\lim_{n\to\infty}[m(E(|f_n-f|\geqslant\sigma)]=0. \tag{21}$$

**注** 证明到此已经得到了一个重要结果：若一个可测函数序列 $\{f_n(x)\}$ 与一个可测函数 $f(x)$ 对一切 $\sigma>0$ 均有(21)式成立，就说 $\{f_n(x)\}$ 按测度收敛于 $f(x)$，因此至此为止，我们已得到了一个结论：

**定理 9（勒贝格定理）** 若可测集 $E$ 上的可测函数序列 $\{f_n(x)\}$ 几乎处处收敛于一个几乎处处取有限值的可测函数 $f(x)$，则 $\{f_n(x)\}$ 必在 $E$ 上按测度收敛于它.

这个定理本身就很有用，下面我们

**继续叶果洛夫定理的证明** 现在取一串下降趋于 0 的正数 $\{\sigma_i\}$，$\sigma_i\to 0$，再取一串正数 $\eta_i$ 使 $\sum\limits_{i=1}^{\infty}\eta_i<+\infty$. 于是由于 $m(R_n(\sigma))\to 0$ 对一切 $\sigma$ 成立，故令 $\sigma=\sigma_i$ 时，一定可以找到 $n_i$ 使 $m(R_{n_i}(\sigma_i))<\eta_i$. 由于 $\sum\limits_{i=1}^{\infty}\eta_i$ 收敛，所以可以找到 $i_0$ 使 $\sum\limits_{i=i_0}^{\infty}\eta_i<\delta$，至此令

$$e=\bigcup_{t=i_0}^{\infty}R_{n_i}(\sigma_i).$$

$e$ 当然是 $E$ 的子集，而 $m(e)\leqslant\sum\limits_{i=i_0}^{\infty}m(R_{n_i}(\sigma_i))<\sum\limits_{i=i_0}^{\infty}\eta_i<\delta$，记

$$E_\delta=E\setminus e$$

今证 $E_\delta$ 即适合所求.

当然

$$m(E_\delta)=m(E)-m(e)>m(E)-\delta.$$

在 $E_\delta$ 上，$f_n(x)$ 必一致收敛于 $f(x)$，事实上任给 $\varepsilon>0$ 总可以找到一个 $i$，适合 $i\geqslant i_0$，$\sigma_i<\varepsilon$，对 $E_\delta$ 中的 $x$，因为 $E_\delta=E\setminus e$，故 $x\notin e$，当然 $x\notin R_{n_i}(\sigma_i)$. 由 $R_{n_i}(\sigma_i)$ 之定义可知，当 $n\geqslant n_i$ 时 $x\notin E(|f_n-f|\geqslant\sigma_i)$，亦即

$$|f_n(x)-f(x)|<\sigma_i<\varepsilon \qquad n\geqslant n_i.$$

因为 $n_i$ 与 $x$ 无关，故定理证毕.

这个定理本身固然重要，它的证法，即分析集合 $\{x: f_n(x)\nrightarrow f(x)\}$ 之构造这个方法更是很重要的.

可测函数在勒贝格积分理论中的地位很像连续函数在黎曼积分理论中的地位. 它当然是连续函数的推广，但其实又不相差太远. 下面我们给出一个有关可测函数的构造的重要定理，但其证明略去.

**定理 10**（鲁金（N.N.Luzin）定理） 若 $f(x)$ 是可测集 $E$ 上的可测函数，则对任一 $\delta>0$，必可找到一个在 $E$ 上连续的函数 $\varphi(x)$，使

$$mE(f\neq\varphi)<\delta,$$

而且若 $|f(x)|\leqslant K$，也有 $|\varphi(x)|\leqslant K$.

**4. 勒贝格积分之定义与基本性质** 有了关于测度与可测函数的准备，就可以着手建立可测集 $E$ 上的可测函数的勒贝格积分理论了，我们首先设 $E$ 是一维区间 $I$.

建立这个理论分成三步，首先考虑有界的可测函数 $f(x)$ 之积分；其次考虑非负（但不一定有界）的可测函数之积分；最后考虑一般情况.

在开始讨论积分理论前，我们先作一个约定，即把 $+\infty$ 与 $-\infty$ 作为两个数来对待，这样做的必要性可以从讨论可测函数序列 $\{f_n(x)\}$ 几乎处处收敛于 $f(x)$ 中看到. 允许 $\{f_n(x)\}$ 在某点不收敛. 但如果 $f_n(x_0)\to+\infty$ 我们最好规定 $f(x_0)=+\infty$，这样可以和 $f_n(x_0)$ 收敛一样来对待. 读者会记得，在第一章我们就指出极限理论实际上建立在潜无限的基础上而不承认实无限，到19世纪下半叶有了以上这些积分理论等等以后，集合论也建立起来了. 这时实无限又一次摆在数学面前，而且成果丰硕，成了整个数学的基础的一部分. 对此我们没有篇幅来介绍了. 在这里我们处理 $\pm\infty$ 的方法只是一种方便的约定罢了. 我们约定

$$\pm\infty+a=\pm\infty \quad (a\text{ 为有限数});$$

$$k\cdot(\pm\infty)=\begin{cases}\pm\infty, & k>0,\\ \mp\infty, & k<0,\end{cases}$$

$$0(\pm\infty)=0;$$

$$(\pm\infty)+(\pm\infty)=\pm\infty;$$

$$(\pm\infty)\cdot(\pm\infty)=+\infty;$$

$$|\pm\infty|=+\infty;$$

$$\frac{a}{\pm\infty}=0 \quad (a\text{ 为有限数}),$$

但是 $(\pm\infty)-(\pm\infty)$，$\dfrac{a}{0}$ 都认为没有意义.

现在设 $f(x)$ 是区间 $A$ 上的有界可测函数，而 $m, M$ 是其下、上确界：$m=\inf\limits_A f(x)$，$M=\sup\limits_A f(x)$，作 $f(x)$ 之值域 $[m,M]$ 的一个分划

$$P: m=y_0<y_1<\cdots<y_n=M \tag{22}$$

并令

$$E_k=E(y_{k-1}\leqslant f(x)<y_k),\ k=1,2,\cdots,n.$$

于是由 $f(x)$之可测性的定义，$E_k$ 都是可测集. 依照定义黎曼积分的方法，我们定义分划 $P$ 的下和 $s_P$ 与上和 $S_P$

$$s_P = \sum_{k=1}^{n} y_{k-1} m(E_k), \quad S_P = \sum_{k=1}^{n} y_k m(E_k). \tag{23}$$

本来，当 $f(x)$不连续时，它在一个小区间上的振动可能很剧烈，使达布和不易确定，现在 $f(x)$虽然只是可测的，我们却是对值域作分划，避免了振动剧烈带来的困难. 但是自变量 $x$ 之相应集合的构造却变得很复杂了. 这一点我们用勒贝格测度来代替若尔当测度来加以克服，于是有了勒贝格的上、下和(23). 和达布上、下和一样，易证当分划加细时上和不增而下的不减，且任一下和必不大于任一上和，因此$\{s_P\}$有上确界 $\underline{I}$，$\{S_P\}$有下确界 $\overline{I}$，它们都是有限数：

$$\underline{I} \leqslant \overline{I} < +\infty.$$

$\underline{I}$ 与 $\overline{I}$ 分别称为 $f(x)$在 $E$ 上的下与上勒贝格积分.

**定义 4**　对定义在区间 $A$ 上的有界可测函数 $f(x)$，若 $\underline{I} = \overline{I}$，就称 $f(x)$在 $A$ 上勒贝格可积 $\underline{I}$ 与 $\overline{I}$ 之公共值 $I$ 称为 $f(x)$在 $A$ 上的勒贝格积分，记作 $\int_A f(x)\mathrm{d}x$ 或 $(L)\int_A f(x)\mathrm{d}x$.

采用后一记号是为了与黎曼积分区别：黎曼积分相应地有时记为 $(R)\int_A f(x)\mathrm{d}x$. 在不会发生误会时就略去记号 $R$ 或 $L$.

**定理 11**　区间 $A$ 上的有界可测函数必为勒贝格可积.

**证**　对某一分划 $P$

$$S_P - s_P = \sum_{k=1}^{n} (y_k - y_{k-1}) m(E_k),$$

当分划足够细使$|y_k - y_{k-1}| < \varepsilon / m(I)$时易得$|S_P - s_P| < \varepsilon$，$\varepsilon$ 是任意小数. 因此 $\underline{I} = \overline{I}$.

举一个例子. 令 $E \subset A$ 是一个可测集，$\chi_E(x)$是其特征函数，容易看到 $E$ 的可测性与 $\chi_E(x)$的可测性是等价的，而且，由定义即得

$$\int_A \chi_E(x)\mathrm{d}x = m(E),$$

就是说，$\chi_E(X)$是可积的(以下如无误会，勒贝格可积就简单地说是可积).

利用 $\chi_E(x)$之可测性与集合 $E$ 之可测性的关系就可以定义可测集 $E$ 上的积分为

$$\int_E f(x)\mathrm{d}x = \int_A \chi_E(x) f(x)\mathrm{d}x. \tag{24}$$

这是因为 $\chi_E(x)f(x)$一方面是有界的，另一方面

$$E(y_\nu \leqslant \chi_E(x) f(x) < y_{\nu+1}) = \begin{cases} [E(y_\nu \leqslant f(x) < y_{\nu+1}) \cap E] & 0 \notin [y_\nu, y_{\nu+1}], \\ [E(y_\nu \leqslant f(x) < y_{\nu+1}) \cap E] \cup (A \setminus E) & 0 \in [y_\nu, y_{\nu+1}], \end{cases}$$

所以 $E(y_\nu\leqslant\chi_E(x)f(x)<y_{\nu+1})$ 又是可测的.

有了勒贝格积分的定义后,我们立刻由定义可知上节(4)式中定义的狄利克雷函数

$$\chi(x)=\begin{cases}1,\ x\text{ 为有理数},\\0,\ x\text{ 为无理数}\end{cases}$$

是勒贝格可积的,而且

$$(L)\int_0^1\chi(x)\mathrm{d}x=0.$$

所以,勒贝格积分确实比黎曼积分更广泛一些.

下面我们进一步考虑非负可测(但不一定有界)函数 $f(x)$ 的勒贝格积分.取 $N$ 为任意大的正整数,并定义

$$[f(x)]_N=\begin{cases}f(x),\text{若 }0\leqslant f(x)<N,\\N,\text{若 }N\leqslant f(x).\end{cases}\tag{25}$$

于是 $[f(x)]_N$ 是有界可测的,从而 $(L)\int_A[f(x)]_N\mathrm{d}x$ 有意义,很容易看到,当 $N$ 增加时 $\{(L)\int_A[f(x)]_N\mathrm{d}x\}$ 不会下降(读者不妨自行证明一下,这就会看到有些对于黎曼积分似乎自明的性质现在证起来却要费一些口舌),因此有两个可能性:一是上述序列有界,从而有极限,我们就定义

$$(L)\int_A f(x)\mathrm{d}x=\lim_{N\to\infty}\int_A[f(x)]_N\mathrm{d}x,\tag{26}$$

并说 $f(x)$ 是勒贝格可积的.二是上述序列无界,有时我们也说这时 $(L)\int_A f(x)\mathrm{d}x=+\infty$.

上面我们是取 $N$ 为正整数的,因此 $\{(L)\int_A[f(x)]_N\mathrm{d}x\}$ 确实是一个序列,但是容易看到取 $N$ 为任意上升趋向 $+\infty$ 的数也是可以的.

最后我们来看一般的可测函数 $f(x)$.我们定义

$$f_+(x)=\begin{cases}f(x),&f(x)\geqslant0,\\0,&f(x)<0;\end{cases}$$

$$f_-(x)=\begin{cases}0,&f(x)\geqslant0,\\-f(x),&f(x)<0;\end{cases}$$

于是 $f_\pm(x)$ 都是非负的,而且容易看到它们都是可测的.而且

$$f(x)=f_+(x)-f_-(x);\ |f(x)|=f_+(x)+f_-(x).$$

我们规定,若 $f_\pm(x)$ 均为 $(L)$ 可积的,则定义 $f(x)$ 为 $L$ 可积,而且定义

$$(L)\int_f(x)\mathrm{d}x=(L)\int_A f_+(x)\mathrm{d}x-(L)\int_A f_-(x)\mathrm{d}x.\tag{27}$$

就是说,我们不但不允许 $(L)\int_A f(x)\mathrm{d}x=a\pm\infty$ 的情形出现,尤其不允许 $(L)\int_A f(x)\mathrm{d}x=(+\infty)-(+\infty)$ 的情形出现,因此我们有很重要的.

**定理 12** 可测函数 $f(x)$ 为可积当且仅当 $|f(x)|$ 为勒贝格可积.

证 若 $f(x)$ 为勒贝格可积，则 $f_{\pm}(x)$ 均为可积，即 $(L)\int_A [f_{\pm}(x)]_N \mathrm{d}x$ 分别收敛于 $(L)\int_A f_{\pm}(x)\mathrm{d}x$，从而

$$(L)\int_A |f(x)|\mathrm{d}x=(L)\int_A f_+(x)\mathrm{d}x+(L)\int_A f_-(x)\mathrm{d}x$$

有意义，而 $|f(x)|$ 为勒贝格可积.

反过来，若 $|f(x)|=f_+(x)+f_-(x)$ 可积，则因 $f(x)$ 可测，故有 $A=E(f(x)\geqslant 0)\cup E(f(x)<0)=A_1\cup A_2$，而 $A_1,A_2$ 均为可测集，而

$$(L)\int_A |f(x)|\mathrm{d}x=(L)\int_{A_1}|f(x)|\mathrm{d}x+(L)\int_{A_2}|f(x)|\mathrm{d}x,$$

这两个积分都是有意义的，从而

$$\int_A \chi_{A_1}(x)f(x)\mathrm{d}x=(L)\int_{A_1}f(x)\mathrm{d}x=(L)\int_A f_+(x)\mathrm{d}x,$$

$$\int_A \chi_{A_2}(x)|f(x)|\mathrm{d}x=-(L)\int_{A_2}|f(x)|\mathrm{d}x=-(L)\int_A f_-(x)\mathrm{d}x$$

都是有意义的，因此由

$$(L)\int_A f(x)\mathrm{d}x=(L)\int_{A_1}f(x)\mathrm{d}x+(L)\int_{A_2}f(x)\mathrm{d}x$$

知 $f(x)$ 在 $A$ 上可积.

不过要注意，上面的证明中我们又用了两个对于黎曼积分似乎自明的性质：若将积分区域分成两块，则一个积分成为两小块上积分之和(积分的可加性)，还有函数和的积分等于函数积分之和(积分的线性).但是对于勒贝格积分，这两个性质下面都将要重新证明.否则定理 12 的证明中有了很大的漏洞.这个证明我们放在下面，现在我们则要指出黎曼积分与勒贝格积分的另一个重要区别在于：考虑无界函数的积分即所谓反常积分时，勒贝格积分与黎曼积分实际上都是用 $[f(x)]_N$ 的积分之极限来处理这个问题，不过黎曼积分并不区别 $f_+(x)$ 与 $f_-(x)$ 而是合并处理，因而在出现 $(+\infty)-(+\infty)$ 的情况下可能有正负抵消而使黎曼积分存在.因此，有些反常积分使 $\int_A f_{\pm}(x)\mathrm{d}x$ 都是有限数而得到绝对收敛性；另一些则是虽然 $\int_A f_{\pm}(x)\mathrm{d}x$ 都是 $+\infty$，但可以用正负抵消使得 $\int_A f(x)\mathrm{d}x$ 有意义即是条件收敛.勒贝格积分则只允许绝对收敛而无所谓条件收敛.例如考虑 $2x\sin\frac{1}{x^2}-\frac{2}{x}\cos\frac{1}{x^2}$ 在 $[0,1]$ 上的积分，这个函数在 $x=0$ 附近是无界的，但是是可测的.其黎曼积分可计算如下：因为 $2x\sin\frac{1}{x^2}-\frac{2}{x}\cos\frac{1}{x^2}=\frac{\mathrm{d}}{\mathrm{d}x}\left(x^2\sin\frac{1}{x^2}\right)$，所以

$$\begin{aligned}(R)\int_0^1\left(2x\sin\frac{1}{x^2}-\frac{2}{x}\cos\frac{1}{x^2}\right)\mathrm{d}x&=\lim_{\varepsilon\to 0}(R)\int_\varepsilon^1\frac{\mathrm{d}}{\mathrm{d}x}\left(x^2\sin\frac{1}{x^2}\right)\mathrm{d}x\\&=\lim_{\varepsilon\to 0}\left(x^2\sin\frac{1}{x^2}\right)\Big|_\varepsilon^1=\sin 1.\end{aligned}$$

但是，因为在$[(2n+\frac{1}{3})\pi]^{-\frac{1}{2}}\leqslant x\leqslant[(2n-\frac{1}{3})\pi]^{-\frac{1}{2}}$时

$$\left|2x\sin\frac{1}{x^2}-\frac{2}{x}\cos\frac{1}{x^2}\right|\geqslant\left|\frac{2}{x}\cos\frac{1}{x^2}\right|-2x\geqslant\frac{1}{x}-2x,$$

由此易见

$$\int_0^1[2x\sin\frac{1}{x^2}-\frac{1}{2}\cos\frac{1}{x^2}]_N\mathrm{d}x>A\ln N,$$

这里 $A$ 是一个适当的正数.从而$(R)\int_0^1\left|2x\sin\frac{1}{x^2}-\frac{1}{x}\cos\frac{1}{x^2}\right|\mathrm{d}x=+\infty$.所以这个函数在黎曼意义下条件收敛但非绝对收敛，而在勒贝格意义下则是不可积的.

下面我们讨论勒贝格积分的性质.读者可能会怀疑，何以在黎曼积分情况下似乎不言自明的性质，现在要花那么多力量去证明？其部分原因在于黎曼积分的积分和的每一项对被积函数都是线性的，而勒贝格积分则不然.这一点将在下面说明，而且正好表明了两种积分确是很不相同的.我们以下只讨论非负可测函数，因为一般的可测(而不一定有界)的函数可以分为两个非负可测函数之差来处理.证明完了这几个性质以后，定理 12 才最终得证.我们把这些性质分为几组最基本的性质来讨论.

(i) **正性** 若 $f(x)$ 在 $E$ 上可积而且 $f(x)\geqslant0$，则

$$(L)\int_E f(x)\mathrm{d}x\geqslant0. \tag{28}$$

反之，若上式中的等号对 $f(x)\geqslant0$ 成立，则 $f(x)=0$ (几乎处处).

**证** 考虑$[f(x)]_N$，并将$[0,N]$分划为

$$P: y_0=0<y_1<\cdots<y_l=N$$

作$\int_E[f(x)]_N\mathrm{d}x$之下和，有

$$s_P=\sum_{\nu=1}^{l-1}y_\nu m(E(y_\nu\leqslant y<y_{\nu+1}))\geqslant0.$$

所以$\sup_P\{s_P\}=(L)\int_A[f(x)]_N\mathrm{d}x\geqslant0$，令 $N\to\infty$ 即得(28).

反之，设 $f(x)\geqslant0$ 而且$(L)\int_E f(x)\mathrm{d}x=0$，作$[f(x)]_N$之值域的划分

$$P:\quad y_0=0<y_1=\frac{1}{n}<\cdots<y_l=N,$$

则因

$$0=(L)\int_E[f(x)]_N\mathrm{d}x\geqslant\sum_{\nu=1}^{l-1}y_\nu m(E(y_\nu\leqslant y\leqslant y_{\nu+1}))+Nm(E(N\leqslant y)),$$

以及 $y_0=0$，而 $\nu\geqslant1$ 时，$y_\nu>0$，$N>0$，所以 $\nu\geqslant1$ 时 $m(E(y_\nu\leqslant y\leqslant\nu_{\nu+1}$

从而 $f(x)$在 $E$ 上几乎处处为 0.

我们要提醒一下,对于黎曼积分,这个结果虽然是对的,却不太容易证明.

用同样的方法可以证明,若在 $E$ 上有界可测的函数 $f(x)$适合不等式 $\alpha\leqslant f(x)\leqslant\beta$,则有中值定理

$$\alpha m(E)\leqslant(L)\int_E f(x)\mathrm{d}x\leqslant\beta m(E). \tag{29}$$

(ii) **可加性** 如果固定 $f(x)$为 $A$ 上的可积函数,则对 $A$ 的每个可测子集 $E$,可以定义 $(L)\int_E f(x)\mathrm{d}x$. 这样,勒贝格积分是一个集合函数,即自变量为一集合(即可测集 $E$)的函数. 和勒贝格测度(测度也是集合函数)一样,这里也有可加性. 具体说来,我们有:若 $f(x)$在 $E$ 上可积,而且 $E=\bigcup\limits_{k=1}^{\infty}E_k$, $E_i\cap E_j=\varnothing$,若 $i\neq j$,则有

$$(L)\int_E f(x)\mathrm{d}x=\sum_{k=1}^{\infty}(L)\int_{E_k}f(x)\mathrm{d}x. \tag{30}$$

我们只需对 $f_+$ 证明上式. 定义$[f]_N$,它是有界的. 如果我们已对有界可测函数证明了上式,则有

$$(L)\int_E[f(x)]_N\mathrm{d}x=\sum_{k=1}^{\infty}(L)\int_{E_k}[f(x)]_N\mathrm{d}x\leqslant\sum_{k=1}^{\infty}(L)\int_{E_k}f(x)\mathrm{d}x.$$

现在令 $N\to\infty$,如果只有有限多个 $E_k$,则只要逐项求极限即得(30),若有无限多个 $E_k$,则由上式左方取 $N\to\infty$时的极限,只能得到一个不等式:

$$(L)\int_E f(x)\mathrm{d}x\leqslant\sum_{k=1}^{\infty}(L)\int_{E_k}f(x)\mathrm{d}x. \tag{31}$$

但对任意的正整数 $K$ 有

$$(L)\int_E[f(x)]_N\mathrm{d}x\geqslant\sum_{k=1}^{K}\int_{E_k}[f(x)]_N\mathrm{d}x.$$

先令 $N\to\infty$,有$(L)\int_E f(x)\mathrm{d}x\geqslant\sum\limits_{k=1}^{K}(L)\int_{E_k}f(x)\mathrm{d}x$,再令 $K\to\infty$又得到相反方向的不等式

$$(L)\int_E f(x)\mathrm{d}x\geqslant\sum_{k=1}^{\infty}(L)\int_{E_k}f(x)\mathrm{d}x. \tag{32}$$

合并(31)与(32)即得(30). 而且由(32)的证明还知道需假设此式右方的无穷级数为收敛才行. 但是如果没有这一点,(30)左、右双方将均为 $+\infty$,这时(30)是平凡不足道的.

至此,余下的仅需对有界可测的 $f(x)$证明(30)成立,证明如下:

先看 $E=E_1\cup E_2$ 且 $E_1\cap E_2=\varnothing$的情况. 于是

$$\begin{aligned}E(y_\nu\leqslant f(x)<y_{\nu+1})&=[E(y_\nu\leqslant f(x)<y_{\nu+1})\cap E_1]\\&\cup[E(y_\nu\leqslant f(x)<y_{\nu+1})\cap E_2]=E_\nu^{(1)}\cup E_\nu^{(2)}\end{aligned}$$

而 $m(E(y_\nu\leqslant f(x)<y_{\nu+1}))=m(E_\nu^{(1)})+m(E_\nu^{(2)})$. 作$\int_E f(x)\mathrm{d}x$ 之下和

$$s_P=s_P^{(1)}+s_P^{(2)}=\sum_\nu[y_\nu m(E_\nu^{(1)})+y_\nu m(E_\nu^{(2)})].$$

先在右方取上确界,然而再在左方取上确界.因为 $f(x)$ 在 $E$,$E_1$ 与 $E_2$ 上均为可积即得

$$(L)\int_E f(x)\mathrm{d}x \leqslant (L)\int_{E_1} f(x)\mathrm{d}x + (L)\int_{E_2} f(x)\mathrm{d}x.$$

再作上和与它们的下确界,又可得到相反的不等式.因此在这个情况下(30)成立.

对于无限多个 $E_k$ 的情况,令

$$S_N = \bigcup_{k=1}^{N} E_k,\quad R_N = E \setminus S_N,$$

于是 $E$ 是 $\{S_N\}$ 的外极限,从而 $m(E) = \lim\limits_{N\to\infty} m(E_N)$,$m(R_N) = m(E) - m(S_N) < \varepsilon$,这里 $\varepsilon$ 是任意给定的正数,而 $N$ 足够大.至此应用上面 $E = E_1 \cup E_2$ 的推理,可知

$$(L)\int_E f(x)\mathrm{d}x = (L)\int_{S_N} f(x)\mathrm{d}x + (L)\int_{R_N} f(x)\mathrm{d}x.$$

对第二项,注意到 $f(x)$ 是有界的:$|f(x)| \leqslant M$,由中值定理(29)以及 $m(R_N) \to 0$,即知第二项趋于0.对第一项再用 $E = E_1 \cup E_2$ 之推理,有

$$(L)\int_{S_N} f(x)\mathrm{d}x = \sum_{k=1}^{N} (L)\int_{E_k} f(x)\mathrm{d}x.$$

令 $N \to \infty$ 即得(30).

(iii) **线性性质** 这里包括两个性质,首先是

$$(L)\int_E [f(x) + g(x)]\mathrm{d}x = (L)\int_E f(x)\mathrm{d}x + (L)\int_E g(x)\mathrm{d}x. \tag{33}$$

我们的证明又要分几步进行.

先设 $f(x)$ 与 $g(x)$ 均为非负,记 $\Phi(x) = f(x) + g(x)$,有

$$[\Phi(x)]_N \leqslant [f(x)]_N + [g(x)]_N \leqslant [f(x) + g(x)]_{2N}.$$

如果对有界可测函数勒贝格积分的线性得证,则由 $\{[f(x)]_N + [g(x)]_N\} - [\Phi(x)_N] \geqslant 0$,利用正性有

$$\int_E \{[f(x)]_N + [g(x)]_N\}\mathrm{d}x - \int_E [\Phi(x)]_N\mathrm{d}x \geqslant 0.$$

对第一项再一次应用有界可测函数的勒贝格积分之线性,有

$$\int_E [\Phi(x)]_N\mathrm{d}x \leqslant \int_E [f(x)]_N\mathrm{d}x + \int_E [g(x)]_N\mathrm{d}x.$$

同理

$$\int_E [f(x)]_N\mathrm{d}x + \int_E [g(x)]_N\mathrm{d}x \leqslant \int_E [\Phi(x)]_{2N}\mathrm{d}x.$$

令 $N \to \infty$,即得(33)

再看 $f$ 与 $g$ 有不同符号,例如 $f(x) \geqslant 0$,$g(x) < 0$ 的情况.先考查集合 $\{x: \Phi(x) \geqslant 0$

下取出.

最一般的情况则是

$$E=E(f(x)\geqslant 0,g(x)>0)\cup E(f(x)<0,g(x)>0)\cup E(f(x)\geqslant 0,g(x)\leqslant 0)$$
$$\cup E(f(x)<0,g(x)\leqslant 0)$$

于是可以看见,线性性质之证明归结为 $f(x)$,$g(x)$均为有界可测——这一点我们马上来证明.以及负号可以由积分号下提出——这一点属于线性性质的另一部分.

设

$$a<f(x)<b,\quad A<g(x)<B.$$

作$[a,b]$与$[A,B]$的分划

$$P: a=y_0<y_1<\cdots<y_n=b,$$

$$Q: A=\bar{y}_0<\bar{y}_1<\cdots<\bar{y}_n=B.$$

令 $e_k=E(y_k\leqslant f(x)<y_{k+1})$,$E_j=E(\bar{y}_j\leqslant g(x)<\bar{y}_{j+1})$. $T_{kj}=e_k\cap E_j$,
于是有

$$E=\bigcup_{k,j}T_{kj},$$

而且 $T_{kj}\cap T_{k'j'}=\varnothing$若$(k,j)\neq(k',j')$.于是利用可加性,有

$$\int_E[f(x)+g(x)]\mathrm{d}x=\sum_{k,j}\int_{T_{kj}}[f(x)+g(x)]\mathrm{d}x.$$

但在 $T_{kj}$上

$$y_k+\bar{y}_j\leqslant f(x)+g(x)<y_{k+1}+\bar{y}_{j+1}.$$

对上式应用积分中值定理并求和,有

$$\sum_{k,j}(y_k+\bar{y}_j)m(T_{kj})\leqslant\int_E[f(x)+y(x)]\mathrm{d}x\leqslant\sum_{k,j}(y_{k+1}+\bar{y}_{j+1})m(T_{kj}).$$

但是由测度的可加性又有例如

$$\sum_{k,j}y_km(T_{kj})=\sum_k y_k\sum_j m(T_{kj})=\sum_k y_km(e_k\cap[\cup E_j])=\sum_k y_km(e_k),$$

这是$(L)\int_{k,j}f(x)\mathrm{d}x$ 的一个下和.同理上式右方是一个上和.求了上、下确界后,利用 $f(x)$与$g(x)$之可积性,即得

$$(L)\int_E[f(x)+g(N)]\mathrm{d}x=(L)\int_E f(x)\mathrm{d}x+(L)\int_E g(x)\mathrm{d}x.$$

线性性质的第二部分是:对任一可积函数 $f(x)$以及任意常数 $c$,$cf(x)$必为可积而且

$$\int_E cf(x)\mathrm{d}x=c\int_E f(x)\mathrm{d}x. \tag{34}$$

证明也要分成几个部分.首先,$c=0$ 时结论是自明的. $c<0$ 时,因为$(-c)>0$,于是由线性性质的第一部分有

$$0=\int_E[cf(x)+(-c)f(x)]\mathrm{d}x=\int_E cf(x)\mathrm{d}x+\int_E(-c)f(x)\mathrm{d}x.$$

若当 $c>0$ 时(34)已得证,则上式右方最后一项应等于 $(-c)\int_E f(x)\mathrm{d}x$. 移项后即有

$$\int_E cf(x)\mathrm{d}x = c\int_E f(x)\mathrm{d}x \qquad (c<0).$$

余下的只是当 $c>0$ 时证明(34)式.

这时仍与前面一样,先把 $f(x)$ 分成 $f_+(x)$ 与 $f_-(x)$;而对非负的 $f(x)$ 又通过 $[f(x)]_N$ 化为有界可测的情况.因此我们只需对有界的 $f(x)$ 与正的 $c$ 证明(34).所以下面我们又设 $A<f(x)<B$,再作 $[A,B]$ 的分划

$$P: A = y_0 < y_1 < \cdots < y_N = B.$$

这就相应于对 $cf(x)$ 之值域作分划

$$cP: cA = \bar{y}_0 < \cdots < \bar{y}_N = cB,$$

这里 $\bar{y}_i = cy_i$,所以

$$E(y_k \leqslant f(x) < y_{k+1}) = E(\bar{y}_k \leqslant cf(x) < \bar{y}_{k+1}).$$

作积分和并求上、下确界即得(34).

(iv) 对积分区域的连续依赖性 设 $f(x)$ 在 $E$ 上可积,$\{E_k\}$ 是一串含于 $E$ 内的可测集:$E_k \subset E$,而且 $m(E_k)=0$,则有

$$\lim_{k\to\infty}\int_{E_k} f(x)\mathrm{d}x = 0.$$

证明如下:不妨设 $f(x)\geqslant 0$,而且取 $N$ 充分大,使

$$\int_E \{f(x)-[f(x)]_N\}\mathrm{d}x < \varepsilon.$$

固定 $N$ 以后,我们有,只要 $k$ 充分大,

$$\int_{E_k} [f(x)]_N \mathrm{d}x \leqslant Nm(E_k) < \varepsilon.$$

于是

$$\begin{aligned}\int_{E_k} f(x)\mathrm{d}x &= \int_{E_k}\{f(x)-[f(x)]_N\}\mathrm{d}x + \int_{E_k}[f(x)]_N\mathrm{d}x.\\ &\leqslant \int_E \{f(x)-[f(x)]_N\}\mathrm{d}x + \int_{E_k}[f(x)]_N\mathrm{d}x < 2\varepsilon.\end{aligned}$$

于是结论得证.

在这个证明中,我们利用了 $f(x)\geqslant 0$, $[f(x)]_N\geqslant 0$, $f(x)-[f(x)]_N\geqslant 0$.

(v) 零测度集的作用.若 $E$ 是零测度集,或 $f(x)$ 几乎处处为 0,都有

$$(L)\int_E f(x)\mathrm{d}x = 0. \tag{35}$$

实际上,$m(E)=0$ 时,(35)左方的积分为 0 可以从定义直接得出.若 $f(x)$ 几乎处处为 0,即 $E_1 = E(f(x)\neq 0)$ 为零测度集,则由积分的可加性有

$$(L)\int_E f(x)\mathrm{d}x = (L)\int_{E_1} f(x)\mathrm{d}x + (L)\int_{E\setminus E_1} f(x)\mathrm{d}x = 0,$$

第二个积分是因为被积函数 $f(x)$ 在 $E\setminus E_1$ 上为 0 而等于 0，第一个积分则由 $m(E_1)=0$ 而为 0.

这个性质告诉我们，在作积分时，积分区域变动一个零测度集或者被积函数在一个零测度集上变动都不会影响积分值.所以零测度集和几乎处处为 0 的函数都是可以忽略的.这样一来，若 $f(x)-g(x)=0$ 几乎处处成立，则 $f(x)$ 和 $g(x)$ 并没有实质上的区别.这种关系是一个等价关系，而几乎处处相等的函数属于同一等价类.在研究积分问题时，我们关心的并不一定是某一特定的函数，而是一个等价类.在讨论有界函数时，这一点尤其重要，可能 $f(x)$ 在一个零测度集 $E_1$ 上无界.删去了 $E_1$，$f(x)$ 在 $E\setminus E_1$ 上可能有界：$\sup\limits_{E\setminus E_1}|f(x)|<\infty$.这样的 $f(x)$ 是本质上有界的.因此应该定义其本质上的"界".但为此就要删去一个适当的 $E_1$，删错了一个 $E_2$，仅是 $m(E_2)=0$，仍有可能 $f(x)$ 在 $E\setminus E_2$ 上无界.所以我们给出

**定义 5** 若 $f(x)$ 在 $E$ 上可测，$E_1\subset E$ 适合 $m(E_1)=0$，我们定义 $f(x)$ 在 $E$ 上的本质上界与下界分别为

$$\text{ess sup} f(x)=\inf\{a\in\mathbf{R}\cup\{+\infty\}, f(x)\leqslant a \text{ 于 } E\setminus E_1 \text{ 上}, m(E_1)=0\}.$$

$$\text{ess inf} f(x)=\sup\{a\in\mathbf{R}\cup\{-\infty\}, f(x)\geqslant a \text{ 于 } E\setminus E_1 \text{ 上}, m(E_1)=0\}.$$

右方是对一切零测度集 $E_1$ 和 $a$ 取上、下确界的.

若 ess sup$f(x)$ 与 ess inf$f(x)$ 均为有限数，则称 $f(x)$ 为本质有界的.本质有界可测函数之集合记为 $L^\infty$.

关于可积函数还有一些简单的性质均可由上面的(i)－(iv)推出，其中有常用的

$$\left|(L)\int_E f(x)\mathrm{d}x\right|\leqslant(L)\int_E|f(x)|\mathrm{d}x, \tag{36}$$

我们就不再证明了.至此我们应该说明，这一些在黎曼积分中十分简单的性质现在证明起来如此费力，是因为黎曼积分和 $\sum f(\xi_k)(x_{k+1}-x_k)$ 一方面保持了线性依赖于 $f$ 的特点，因此黎曼积分的线性性质几乎是自明的.另一方面又是因为黎曼积分是对自变量的区间(而不是一般集合)作最简单的分划，其长度关系是一个最简单的组合学关系式

$$b-a=(b-x_{n-1})+\cdots(x_2-x_1)+(x_1-a).$$

勒贝格积分正因为放弃了这两点而允许对值域作分划——因而积分的线性性质隐晦了；它还允许研究很一般的集合的测度——因而以放弃上述组合学关系式为代价突出了可数可加性.

至此我们要问，这两种积分关系如何？这里我们有基本的

**定理 13** 若可测函数 $f(x)$ 在区间 $A$ 上黎曼可积，则它亦必勒贝格可积，而且

$$(R)\int_A f(x)\mathrm{d}x=(L)\int_A f(x)\mathrm{d}x. \tag{37}$$

**证** 上一节中我们已证明了黎曼可积函数必几乎处处连续，所以除一个零测度集以外 $f(x)$ 是可测的.加上这个零测度集，$f(x)$ 在 $A$ 上是可测的.黎曼可积函数必为有界，因此 $f(x)$ 有界可测，也就是勒贝格可积的.现将 $A$ 分为有限多个子区间：$A=\bigcup\limits_{k=1}^{N}A_k$，则由勒贝格积分的可加性，令 $m_k=\inf\limits_{A_k}f(x)$，$M_k=\sup\limits_{A_k}f(x)$，由勒贝格积分之中值定理有

$$\sum_k m_k m(A_k) \leqslant (L)\int_A f(x)\mathrm{d}x \leqslant \sum_k M_k m(A_k),$$

但是两端两项是黎曼积分的上下和. 故若 $A_k$ 的直径趋于 0, 由假设 $f(x)$ 为黎曼可积即得(37). 证毕.

这个定理很重要: 当我们要计算 $(L)\int_A f(x)\mathrm{d}x$ 时, 只要 $f(x)$ 黎曼可积, 就可以把它当作黎曼积分来算; 当我们要研究 $(R)\int_A f(x)\mathrm{d}x$ 的由下文诸定理表示的性质时, 就可以把它当作勒贝格积分并应用下面这些强有力的定理. 所以, 下面若无特别需要, 我们就只写 $\int_A f(x)\mathrm{d}x$ 而不注明 $(R)$ 或 $(L)$.

我们讲的"这些强有力的定理"就是涉及黎曼积分局限性的一些定理. 首先要证明基本的

**定理 14** (**勒贝格控制收敛定理**) 设可测函数序列 $\{f_n(x)\}$ 在 $A$ 上对 $n$ 一致有界, 即 $|f_n(x)| \leqslant M$, 又在 $A$ 上几乎处处有 $\lim\limits_{n\to\infty} f_n(x) = f(x)$, 则 $f(x)$ 在 $A$ 上也是勒贝格可积的, 而且可以在积分号下求极限:

$$\lim_{n\to\infty}\int_A f_n(x)\mathrm{d}x = \int_A f(x)\mathrm{d}x. \tag{38}$$

**注 1** $\{f_n(x)\}$ 一致有界可以改成 $|f_n(x)| \leqslant \Phi(x)$ 而 $\Phi(x)$ 勒贝格可积.

**注 2** 令 $E = \{x: \lim\limits_{n\to\infty} f_n(x) \neq f(x)\}$, 则 $m(E) = 0$. 在 $E$ 上把 $f_n(x)$ 与 $f(x)$ 都改成 0, 不会影响积分而定理条件可改为 $\lim\limits_{n\to\infty} f_n(x) = f(x)$ 处处成立.

**定理的证明** 可测函数 $f_n(x)$ 之极限函数 $f(x)$ 必可测, 又因 $|f(x)| = |\lim\limits_{n\to\infty} f_n(x)| \leqslant M$, 故 $f(x)$ 为有界可测从而勒贝格可积.

对任意 $\varepsilon$, 由叶果洛夫定理必可找到 $A$ 的子集 $E$, 使 $E$ 上 $f_n(x)$ 一致收敛于 $f(x)$, 而且 $m(A \setminus E) \leqslant \dfrac{\varepsilon}{4M}$, 于是

$$\begin{aligned}\left|\int_A [f_n(x) - f(x)]\mathrm{d}x\right| &\leqslant \int_E |f_n(x) - f(x)|\mathrm{d}x + \int_{A\setminus E} |f_n(x) - f(x)|\mathrm{d}x \\ &\leqslant \int_E |f_n(x) - f(x)|\mathrm{d}x + \frac{\varepsilon}{2}.\end{aligned}$$

由 $f_n(x)$ 在 $E$ 上一致收敛于 $f(x)$, 故有 $N(\varepsilon)$ 存在, 使 $n > N(\varepsilon)$ 时 $|f_n(x) - f(x)| < \dfrac{\varepsilon}{2m(A)}$, 代入上式即知当 $n > N($

分.积分区域 $E$ 为有限测度的可测集而 $f(x)$ 在 $E$ 上不一定有界的情况前已讨论过了:我们用 $\int_E [f(x)]_N \mathrm{d}x$ 之极限作为 $\int_E f(x)\mathrm{d}x$ 之定义,这一作法其实与黎曼积分情况是一致的;而另一类"反常积分"即 $E$ 不一定具有限测度的集合的处理方法也类似:我们还是先设 $f(x)\geqslant 0$,因为一般情况下可以把 $f(x)$ 写成 $f(x)=f_+(x)-f_-(x)$,并分别讨论 $\int_E f_+(x)\mathrm{d}x$ 与 $\int_E f_-(x)\mathrm{d}x$.当二者均有定义(即均收敛)时,我们定义 $\int_E f(x)\mathrm{d}x=\int_E f_+(x)\mathrm{d}x-\int_E f_-(x)\mathrm{d}x$.这样做即知 $\int_E |f(x)|\mathrm{d}x=\int_E f_+(x)\mathrm{d}x+\int_E f_-(x)\mathrm{d}x$ 也有定义(即也收敛).所以,和 $E$ 为有限测度可测集的情况一样,勒贝格可积函数一定是绝对可积的.我们现在限制讨论这样的 $E$:令 $\{\Omega_k\}$ 是有界闭域(即紧区域)的"穷竭的上升"序列.这句话的意思是,首先,$\{\Omega_k\}$ 是上升序列,$\Omega_1\subset\Omega_2\subset\cdots$;其次 $E$ 的任一紧子集必包含在某一 $\Omega_{k_0}$ 中,所以 $\bigcup\limits_{n=1}^{\infty}\Omega_k=E$;其次 $E_k=E\cap\Omega_k$ 具有有限测度,而 $E=\lim\limits_{k\to\infty}E_k=\bigcup\limits_{k=1}^{\infty}E_k$,即为 $\{E_k\}$ 之外极限.(当然,我们可以利用区间 $I_k$——其中心在 $x=0$,边长为 $k$,来作 $E_k$,也可以用 $k$ 为半径,而球心在 $x=0$ 的球 $B_k$ 来作 $E_k$:$E_k=E\cap B_k$,对下面的结果不会有影响.)如果 $f(x)$ 在 $E_k$ 上可积,而且 $\lim\limits_{k\to\infty}\int_{E_k} f(x)\mathrm{d}x$ 存在(因为已设 $f(x)\geqslant 0$,所以这个极限若不存在,必有 $\int_{E_k} f(x)\mathrm{d}x\to+\infty$),就说 $f(x)$ 在 $E$ 上勒贝格可积,而且定义 $(L)\int_E f(x)\mathrm{d}x=\lim\limits_{k\to\infty}(L)\int_{E_k} f(x)\mathrm{d}x$.(我们再提一下,如果不用 $I_k$ 而用 $B_k$ 来作 $E_k$,则关于 $f(x)$ 之可积性与积分值只会得到相同结果).勒贝格积分一定是绝对可积的,这是一个重要性质,例如 $\int_{-\infty}^{\infty}\frac{\sin x}{x}\mathrm{d}x$ 作为黎曼积分是存在的——条件收敛,但作为勒贝格积分则是不存在的.

前面讲的关于勒贝格积分的性质大部分在此仍成立,当然也有例外,如中值定理 $\alpha m(E)\leqslant\int_E f(x)\mathrm{d}x\leqslant\beta m(E)$,因为现在 $m(E)=+\infty$,这一切读者都容易判断.重要的是要指出,勒贝格控制收敛定理仍成立,但 $|f(x)|\leqslant M$ 这个条件要代以 $|f(x)|\leqslant\Phi(x)$,这里 $\Phi(x)$ 在 $E$ 上可积.

现在我们开始讨论含参变量的积分.设有函数 $f(x,t)$,这里 $x=(x_1,\cdots,x_n)\in\mathbf{R}^n$,$t$ 是参数:$t=(t_1,\cdots,t_m)\in U\subset\mathbf{R}^m$,$f(x,t)$ 可以取 $\pm\infty$ 为值.于是我们有

**定理 15** (i) 设对任一固定的 $t\in U$,$f(x,t)$ 对 $x$ 在 $\mathbf{R}^n$ 上可积.

(ii) 对几乎所有 $x\in\mathbf{R}^n$,$f(x,t)$ 作为 $t$ 的函数在 $t_0\in U$ 处连续.

(iii) 存在一个 $\mathbf{R}^n$ 上可积的函数 $\Phi(x)$,使对一切 $t\in U$,对于 $x$ 几乎处处有

$$|f(x,t)|\leqslant\Phi(x),$$

则 $g(t)=\int_{\mathbf{R}^n} f(x,t)\mathrm{d}x$

**证**　我们只要证明对任一串 $t_n \to t_0$，$g(t_n) \to g(t_0)$ 即可. 为此，我们令

$$f_n(x) = f(x, t_n), f_0(x) = f(x, t_0),$$

于是由(i)知 $f_n(x)$ 与 $f_0(x)$ 均在 $\mathbf{R}^n$ 上勒贝格可积. 由(ii)知，对几乎所有 $x$，$\lim\limits_{n\to\infty} f_n(x) = f_0(x)$，由(iii)知 $|f_n(x)| \leqslant \Phi(x)$，所以由勒贝格控制定理

$$\lim_{n\to\infty} g(t_n) = \lim_{n\to\infty} \int_{\mathbf{R}^n} f(x, t_n)\,\mathrm{d}x = \int_{\mathbf{R}^n} f(x, t_0)\,\mathrm{d}x = g(t_0).$$

定理证毕.

**注**　定理中 $x \in \mathbf{R}^n$ 可以改为 $x \in E$，而 $E$ 具有上面讲的性质，为此只需将 $f(x,t)$ 改为 $\chi_E(x) f(x,t)$ 即可.

**定理 16**　(i) 仍设 $f(x,t)$ 对固定的 $t \in U$ 为可积. 但现在设 $U \subset \mathbf{R}^m$ 是一个开集.

(ii) 对几乎所有 $x$，$f(x,t)$ 作为 $t$ 的函数是 $U$ 中的可微函数.

(iii) 存在一个 $\mathbf{R}^n$ 上可积的函数 $\Phi(x)$，使对所有的 $t = (t_1, \cdots, t_m) \in U$，$\left|\dfrac{\partial f(x,t)}{\partial t_i}\right| \leqslant \Phi(x)$ $(i = 1, 2, \cdots, m)$ 对几乎所有 $x \in \mathbf{R}^n$ 成立.

则 $g(t) = \int_{\mathbf{R}^n} f(x,t)\,\mathrm{d}x$ 对 $t$ 在 $U$ 中可微，而且

$$\frac{\partial g(t)}{\partial t_i} = \int_{\mathbf{R}^n} \frac{\partial f(x,t)}{\partial t_i}\,\mathrm{d}x, i = 1, 2, \cdots, m.$$

**证**　我们只需证明对任一个 $t_i$，$\dfrac{\partial g(t)}{\partial t_i}$ 为连续且可以在积分号下求导. 所以，不失一般性不妨设 $U$ 是一维空间中一个开区间 $(t_0 - h, t_0 + h)$，而证明 $g(t)$ 在 $t_0$ 处连续可微. 令 $h_i$ 为一个趋于 0 的实数列，$h_i \neq 0$，而记

$$f_i(x) = \frac{f(x, t_0 + h_i) - f(x, t_0)}{h_i},$$

$$f_0(x) = \frac{\partial f(x, t_0)}{\partial t}.$$

于是由拉格朗日定理

$$f_i = f'(x, t_0 + \theta_i h_i), 0 < \theta_i < 1.$$

仿照定理 15 的证法令 $i \to \infty$ 立即可得如上的结果.

定理 15 与定理 16 比之黎曼积分中的相应定理显然更简单好用. 读

中就见不到.这些思想与方法正是勒贝格积分理论与黎曼积分理论本质的区别.下面我们要讲的单调性问题也是其中之一,它反映了勒贝格积分的"正性"起了何等重要的作用,我们先证明

**定理 17(法图(Fatou)引理)**　设 $E$ 是有限测度集,而可测函数 $f_n(x)\geqslant 0$ 于 $E$ 上,而且几乎处处有 $f_n(x)\to f(x)$,则

$$\int_E f(x)\mathrm{d}x\leqslant\varliminf_{n\to\infty}\int_E f_n(x)\mathrm{d}x. \tag{39}$$

**证**　很容易证明,在 $E$ 上几乎处处有

$$\lim_{n\to\infty}[f_n(x)]_N=[f(x)]_N\leqslant N.$$

故由勒贝格控制收敛定理

$$\lim_{n\to\infty}\int_E[f_n(x)]_N\mathrm{d}x=\int_E[f(x)]_N\mathrm{d}x, \tag{40}$$

但对左方的积分有

$$\int_E[f_n(x)]_N\mathrm{d}x\leqslant\int_E f_n(x)\mathrm{d}x.$$

左方有极限存在,右方则不一定,但右方的下极限一定存在,故

$$\int_E[f(x)]_N\mathrm{d}x\leqslant\varliminf_{n\to\infty}\int_E f_n(x)\mathrm{d}x.$$

代入(40)有再令 $N\to\infty$ 即得(39)式.

对这个定理的理解中要注意,若(39)之右方是有限数,则知 $f(x)$ 在 $E$ 可积.反过来,$f(x)$ 总是非负可测的(定理 8),若它不可积,有 $\int_E f(x)\mathrm{d}x=+\infty$,这时必有 $\varliminf_{n\to\infty}\int_E f_n(x)\mathrm{d}x=+\infty$,因此我们仍认为(39)成立,且其中有等号.

利用法图引理就可以证明重要的

**定理 18(勒维(Beppo Levi)定理)**　令 $\{f_n(x)\}$ 为 $E$ 上的可积函数的上升序列,而且 $\lim_{n\to\infty}f_n(x)=f(x)$,则在以下意义下可以在积分号下取极限

$$\lim_{n\to\infty}\int_E f_n(x)\mathrm{d}x=\int_E\lim_{n\to\infty}f_n(x)\mathrm{d}x=\int_E f(x)\mathrm{d}x: \tag{41}$$

(i) 若式左为有限,则 $f(x)$ 几乎处处有限而且为可积,且等号成立;若式左为 $+\infty$,则 $f(x)$ 不可积,但(41)仍成立.

(ii) 若极限函数 $f(x)$ 可积.则上式两侧均为有限且相等;若极限函数 $f(x)$ 不可积而式左成为 $+\infty$,而(41)式仍成立.

**证** (i) 不妨考虑 $f_n(x)-f_1(x)$,于是可以设 $f_n(x)\geqslant 0$ 由法图引理知,若式左为有限,$f(x)$ 在 $E$ 上可积,由勒贝格控制定理(注意 $|f_n(x)|=f_n(x)\leqslant f(x)$),知(41)成立.式左为 $+\infty$ 时,也由法图引理可知式右也是 $+\infty$,故(41)仍成立.

(ii) 若 $f(x)$ 可积,由 $|f_n(x)|=f_n(x)\leqslant f(x)$ 用勒贝格控制收敛定理仍可得(41)式,若 $f(x)$ 不可积,因 $f(x)\geqslant 0$,故 $\int_E f(x)\mathrm{d}x=+\infty$,由法图引理知 $\varliminf_{n\to\infty}\int_E f_n(x)\mathrm{d}x=+\infty$,从而 $\lim_{n\to\infty}\int_E f_n(x)\mathrm{d}x=+\infty$ 而(41)式仍成立.

这个定理对于下降函数序列也是对的,因为$\{-f_n(x)\}$这时是上升序列.

对于 $E$ 不一定具有有限测度的情况,仿照上面的方法仍可证明这两个定理.

勒维定理常用于正项级数的情况,即将 $f_n(x)$表为 $\sum_{n=1}^{\infty} g_n(x)$的部分和,$f_1(x)=g_1(x)$,$f_n(x)-f_{n-1}(x)=g_n(x)$,这时我们有:若 $g_n(x)\geqslant 0$ 为可积函数,则

$$\int_E \sum_{n=1}^{\infty} g_n(x)\mathrm{d}x = \sum_{n=1}^{\infty}\int_E g_n(x)\mathrm{d}x. \tag{42}$$

把它与上节关于黎曼积分的定理 12 比较一下:对非负黎曼可积的 $g_n(x)$,要把(42)左方的积分改为下积分才行,尽管我们这时有看来很强的迪尼定理(即上节定理 13),它指出若 $f_n(x)\geqslant 0$ 为连续,且在有界闭区间上下降趋于 0,则它必一致趋于 0.问题在于由连续函数到可测函数是一个复杂的过程,上面我们仅举出一个鲁金定理,在以下各章我们还有机会讲到进一步的结果.

**5.勒贝格积分怎样改进黎曼积分的不足之处** 前节末尾我们指出例如在变分问题中我们需要以下类型的定理:若有一串可积函数$\{u_n(x)\}$,而且对任意 $\varepsilon>0$ 都有正整数 $N=N(\varepsilon)$存在,使当 $n,m>N$ 时,$\int_N |u_m(x)-u_n(x)|\mathrm{d}x<\varepsilon$,这时是否有一个可积函数 $u(x)$存在,使$\int_E |u_n(x)-u(x)|\mathrm{d}x<\varepsilon$,亦即在某种意义下$\lim_{n\to\infty}u_n(x)=u(x)$?这样提出问题很像极限理论中的柯西准则:若有一个实数柯西序列$\{a_n\}$,则必有一实数 $a$ 存在,使$\lim_{n\to\infty}a_n=a$.这当然是要证明的.我们在第六章中讲实数理论时要讲到,如果 $a_n$ 全是有理数,则若限制在有理数中,这个极限 $a$ 一般是找不到的,一定要对有理数集合中补充上无理数成为实数集合,才一定能找到作为极限的实数 $a$.但若 $a_n$ 为实数(有理或无理),则一定可找到实数 $a$ 而用不着再补充新种类的数了.这就称为实数的完备性以及有理数系的不完备性.现在的情况很类似:如果规定只考虑黎曼可积函数,则不一定有黎曼可积的极限函数 $u(x)$使得$\lim_{n\to\infty}\int_E |u_n(x)-u(x)|\mathrm{d}x=0$,但若规定 $u_n(x)$是勒贝格可积,则一定有勒贝格可积的 $u(x)$使上式成立.所以这个性质称为勒贝格可积函数空间的完备性.下面我们就要证明它(以后还要证明,一切勒贝格可积函数都可以这样构造出来),这是勒贝格积分理论的一个核心定理.

**定理 19(里斯-费希尔(Riesz-Fisher)定理)** 设 $E$ 为具有限测度的可测集,$\{f_n(x)\}$是 $E$ 上的勒贝格可积函数序列,而且

$$\int_E |f_n(x)-f_m(x)|\mathrm{d}x\to 0,\text{当 } n,m\to\infty\text{时}, \tag{43}$$

则必存在在 $E$ 上勒贝格可积的函数 $f(x)$,使当 $n\to\infty$ 时

$$\int_E |f$$

$$\int_E |f_{n_{k+1}}(x)-f_{n_k}(x)|\mathrm{d}x<2^{-k}.$$

令 $u_k(x)=f_{n_{k+1}}(x)-f_{n_k}(x)$，$u_0(x)=f_{n_1}(x)$，则 $u_k(x)$勒贝格可积，而且

$$\sum_{k=0}^{\infty}\int_E |u_k(x)|\mathrm{d}x<\sum_{k=1}^{\infty}2^{-k}<+\infty.$$

由勒维定理，$\sum\limits_{k=0}^{\infty}|u_k(x)|$几乎处处收敛，其和 $F(x)$也是勒贝格可积的. 进而看级数 $\sum\limits_{k=0}^{\infty}u_k(x)$，它必几乎处处绝对收敛. 令其部分和为 $S_N(x)$，则$\lim\limits_{N\to\infty}S_N(x)=f(x)$几乎处处成立，而且是可测函数，但是

$$|S_N(x)|\leqslant\sum_{k=0}^{N}|u_k(x)|\leqslant\sum_{k=0}^{\infty}|u_k(x)|=F(x).$$

而且$|f(x)-S_N(x)|\to 0$，几乎处处成立，$|f(x)-S_N(x)|\leqslant\sum\limits_{k=N+1}^{\infty}|u_k(x)|\leqslant F(x)$，再用一次勒贝格控制收敛定理有

$$\int_E |f(x)-S_N(x)|\mathrm{d}x=\int_E |f(x)-f_{n_{N+1}}(x)|\mathrm{d}x\to 0,$$

特别是

$$\int_E |f(x)-f_n(x)|\mathrm{d}x\leqslant\int_E |f(x)-f_{n_{N+1}}(x)|\mathrm{d}x$$
$$+\int_E |f_{N_{n+1}}(x)-f_n(x)|\mathrm{d}x\to 0.$$

定理证毕.

如果 $E$ 不一定有有限测度而与定理 15,16 一样，这个定理仍成立.

关于勒贝格积分的基本概念和性质，我们就介绍至此. 数学中还有多种其它积分理论，但是都适合同样的框架：每个积分都是一个泛函，即对某一函数类——可积函数类——的一个元 $f$ 给定一个值$I(f)$. 如果 $I(f)$具有某些性质，则可以利用这些性质来对积分理论作公理化的处理. 例如我们可以要求

(i) $I(f)$是线性泛函，$I(af+bg)=aI(f)+bI(g)$，这里 $a$，$b$ 是实数.

(ii) 正性：$f\geqslant 0$ 时，$I(f)\geqslant 0$.

(iii) 关于单调收敛的连续性：若 $f_n(x)$下降趋于 0，则 $I(f_n)$也下降趋于 0.

黎曼积分和勒贝格积分都具有这些性质，其中的(iii)，对于黎曼积分即是迪尼定理，对于勒贝格积分即是勒维定理. 进一步我们可以称适合它们的泛函就是积分，这样就为积分理论提供了一个公理化框架. 它与勒贝格积分理论不同，它不是以测度理论为基础的. 这种积分有时称为丹尼尔(Daniell 积分)，它是很有用处的.

## § 3　勒贝格积分的初步介绍(续)

**1. 有界变差函数**　上一节我们讨论了函数在勒贝格意义下的可积性以及勒贝格积分的主

要性质，特别是我们证明了里斯—费希尔定理从而说明了勒贝格积分确实在完备性问题上克服了黎曼积分的缺点.我们也看到，之所以能够得到较好的结果是由于我们使用了具有可数可加性的勒贝格测度，这样可以用分划 $f(x)$之值域来代替分划自变量 $x$ 的变化区域即$f(x)$之定义域.当然，这样做是有代价的：我们不得不放弃黎曼积分和对被积函数的线性关系，以及自变量变化区域中的组合关系：$b-a=(b-x_{n-1})+(x_{n-1}-x_{n-2})+\cdots+(x_1-a)$.这样，一些在黎曼积分中几乎是自明的事情，现在证明起来颇多麻烦.本节中我们还要讲两个重要问题：积分在什么意义下是微分的逆运算以及二重积分化为逐次积分的问题.本节的讨论也有上面说的特点，即论证比较细致使人感到繁琐，而结论则似乎并不多于连续函数的黎曼积分中的相应结果，给人的感觉是不太合算.因此，读者完全可以跳过本节到他认为自己比较习惯于这种细密的推理时再回到本节.好在本节的结论读者现在就会认为是不言而喻的.

关于积分与微分互为逆运算的问题，读者都十分习惯于柯西的结果，若 $f(x)$在$[a,b]$上连续，则$\int_a^x f(t)\mathrm{d}t$ 是$f(x)$在$[a,b]$上的原函数，若能再找到一个原函数 $F(x)$，则必有

$$\int_a^b f(x)\mathrm{d}x=F(b)-F(a). \tag{1}$$

到了黎曼的时代，因为不连续函数已进入了人们的视野，达布发现，应该区别关于 $f(x)$的两件事，一是可积，二是有原函数 $F(x)$.(注意，这里原函数是指处处适合 $F'(x)=f(x)$的函数.)上面两件事是互相独立的.这时$\int_a^x f(t)\mathrm{d}t$ 是否也是一个原函数，且(1)成立？

于是自然要问，在勒贝格积分的框架下能得到什么样的结果？是否能使积分与求导恰为逆运算？遗憾的是，这是办不到的.现在我们把问题分成两个：

Ⅰ.若 $f(x)$在$[a,b]$上勒贝格可积，则 $F(x)=\int_a^x f(t)\mathrm{d}t$ 是否$f(x)$的原函数，而(1)成立？

首先，原函数意义要改动一下，因为若在一个零测度集上改变 $f(x)$，则$\int_a^x f(t)\mathrm{d}t$ 不会改变，从而 $F'(x)=f(x)$只能几乎处处成立.所以下面凡讲到原函数均指几乎处处意义 下的原函数.

Ⅱ.$F(x)$要适合什么条件才有某个勒贝格可积的导函数 $f(x)=F'(x)$，而且使 $F(x)=\int_a^x f(t)\mathrm{d}t+C$？

第一个问题答案很清楚.若 $f(x)$为勒贝格可积，则 $F(x)=\int_a^x f(t)\mathrm{d}t$ 一定几乎处处以$f(x)$为导数.这里要利用以下的性质，即不定积分必为有界变差函数.那么，什么是有界变差函数呢？以下我们只限于一元函数，因为多元函数的情况是很不一样的.

设 $f(x)$定在$[a,b]$上，现在作$[a,b]$的一个分划

$$P: a=x_0<x_1<\cdots<x_n=b.$$

作 $V=\sum_{i=0}^{n-1}|f(x_{i-1})-f(x_i)|$，则$\sup_P V$ 称为$f(x)$在$[a,b]$上的全变差.

**定义 1** 若 $V_a^b(f)=\sup_P V<+\infty$，则称 $f(x)$是$[a,b]$上的有界变差函数.

很明显，单调函数 $f(x)$都是有界变差的.因为不失一般性，不妨限于考虑单调不减的$f(x)$.

这时，$|f(x_{i+1})-f(x_i)|=f(x_{i+1})-f(x_i)\geqslant 0$，因此

$$\begin{aligned}V&=|f(x_n)-f(x_{n-1})|+\cdots+|f(x_1)-f(x_0)|\\&=[f(x_n)-f(x_{n-1})]+\cdots+[f(x_1)-f(x_0)]=f(b)-f(a).\end{aligned}$$

所以，$V_a^b(f)=\sup_P V=f(b)-f(a)<+\infty$，从而单调函数 $f(x)$ 具有有界变差. 重要的是其逆在某种程度上也成立，我们有

**定理 1**　$[a,b]$上的有界变差函数 $f(x)$ 必可写为两个单调不减函数之差.

**证**　任取 $c\in[a,b]$，于是得到$[a,b]=[a,c]\cup[c,b]$，当我们作$[a,b]$之任一分划 $P$ 时都不妨把 $c$ 当作一个分点：$c=x_m$，于是

$$\sum_{i=0}^{n-1}\left|f(x_{i+1})-f(x_i)\right|=\sum_{i=0}^{m-1}\left|f(x_{i+1})-f(x_i)\right|+\sum_{i=m}^{n-1}\left|f(x_{i+1})-f(x_i)\right|.\tag{2}$$

现在从左、右双侧来对分划 $P$ 取上确界.

先看右方，固定 $c$ 不动，而得到$[a,c]$的分划 $P_1$，$[c,b]$的分划 $P_2$，因此

$$\begin{aligned}\sum_{i=0}^{m-1}\left|f(x_{i+1})-f(x_2)\right|&\leqslant V_a^c(f),\\ \sum_{i=m}^{n-1}\left|f(x_{i+1})-f(x_i)\right|&\leqslant V_c^b(f).\end{aligned}\tag{3}$$

我们很容易证明，当 $f$ 在$[a,b]$上有有界变差时，它在任一子区间上也有有界变差. 所以上面两个式子右方均为有限数，但是我们并不需要这件事，读者如愿去证明也不难. 以(3)代入(2)之右方即知

$$\sum_{i=0}^{n-1}|f(x_{i+1})-f(x_i)|\leqslant V_a^c(f)+V_c^b(f)$$

此式对任何以 $c$ 作为一个分点的分划$P$ 均成立. 对于不以 $c$ 为分点的分划，在增加一个分点 $c$ 以后，其相应的变差 V 不减，所以上式对任意分划 $P$(不论是否含 $c$ 作为分点)均成立. 取上确界以后即得

$$V_a^b(f)\leqslant V_a^c(f)+V_c^b(f)\tag{4}$$

另一方面，从(2)的左方看应有

$$\sum_{i=0}^{m-1}\left|f(x_{i+1})-f(x_i)\right|+\sum_{i=m}^{n-1}\left|f(x_{i+1})-f(x_i)\right|\leqslant V_a^b(f)$$

作适当分划，总可以使 $\sum\limits_{i=1}^{m-1}|f(x_{i+1})-f(x_i)|\geqslant V_a^c(f)-\varepsilon$，$\varepsilon>0$ 是任意小正数，同样 $\sum\limits_{i=m}^{n-1}\left|f(x_{i+1})-f(x_i)\right|\geqslant V_c^b(f)-\varepsilon$，代入上式再令 $\varepsilon\to 0$，又有

$$V_a^c(f)+V_c^b(f)\leqslant V_a^b(f).\tag{5}$$

比较(4)，(5)两式即有

$$V_a^b(f)=V_a^c(f)+V_c^b(f).\tag{6}$$

现在把 $V_a^b(f)$中的 $b$ 代以变量$x$，则得到一个函数 $V(x)$，$V(x)$显然是不减的，因为若 $x_1<x_2$，则由(6)

$$V(x_2)=V(x_1)+V_{x_1}^{x_2}(f).$$

但是 $V_{x_1}^{x_2}(f)\geqslant 0$，这由全变差之定义自明．考虑差 $W(x)=V(x)-f(x)$ 可以证明这个差 $W(x)$ 也是不减的，这是因为

$$\begin{aligned}W(x_2)-W(x_1)&=V(x_2)-V(x_1)-[f(x_2)-f(x_1)]\\&=V_{x_1}^{x_2}(f)-[f(x_2)-f(x_1)]\geqslant V_{x_1}^{x_2}(f)-|f(x_2)-f(x_1)|.\end{aligned}$$

但是很明显有 $|f(x_2)-f(x_1)|\leqslant V_{x_1}^{x_2}(f)$，因为左方只是 $f(x)$ 在区间 $[x_1,x_2]$ 对一个特殊分划——分点只有 $x_1,x_2$ 两点——的变差，而右方是 $f(x)$ 对一切分划的变差的上确界，所以，

$$W(x_2)\geqslant W(x_1),$$

而 $W(x)$ 是不减的，由于 $f(x)=V(x)-W(x)$，故定理得证．

**推论** 若有界变差函数 $f(x)$ 还是 $[a,b]$ 上的连续函数，则它必可表示为两个连续不减函数之差．

**证** 我们只需证明 $V(x)$ 为连续即可．取 $x_0\in[a,b]$．为方便起见我们来看 $V(x)$ 在 $x_0$ 之右连续性（左连续证明一样）．因为 $f(x)$ 在 $x_0$ 连续，故必为右连续．对任意 $\varepsilon>0$，必可找到 $\delta(\varepsilon)>0$，使若 $x_1\in[x_0,x_0+\delta]$，当有 $|f(x_1)-f(x_0)|<\dfrac{\varepsilon}{2}$．因为 $f(x)$ 在 $[a,b]$ 上为有界变差的，所以在 $[x_0,b]$ 上也有有界变差．因此对任意 $\varepsilon>0$ 都可以找到 $[x_0,b]$ 的一个分划 $P$：$x_0<x_1<\cdots<x_n=b$，而且 $|x_1-x_0|<\delta$，使得

$$\sum_{i=0}^{n-1}|f(x_{i+1})-f(x_i)|>V_{x_0}^{b}(f)-\frac{\varepsilon}{2},$$

或者

$$\begin{aligned}V_{x_0}^{b}(f)&<\sum_{i=0}^{n-1}|f(x_{i+1})-f(x_i)|+\frac{\varepsilon}{2}\\&=|f(x_1)-f(x_0)|+\sum_{i=1}^{n-1}|f(x_{i+1})-f(x_i)|+\frac{\varepsilon}{2}\\&<\varepsilon+\sum_{i=1}^{n-1}|f(x_i\ \cdots\end{aligned}$$

$$\sum_{i=0}^{n-1} |[af_1(x_{i+1})+bf_2(x_{i+1})]-[af_1(x_i)+bf_2(x_i)]|$$
$$\leqslant |a| \sum_{i=0}^{n-1} |f_1(x_{i+1})-f_1(x_i)| + |b| \sum_{i=0}^{n-1} |f_2(x_{i+1})-f_2(x_i)|$$
$$\leqslant |a| \mathrm{V}_\alpha^\beta(f_1) + |b| \mathrm{V}_\alpha^\beta(f_2)$$

即知.

**注**　也很容易证明 $f_1(x)\cdot f_2(x)$，$f_1(x)/f_2(x)$(若 $f_2(x)\neq 0$)也都是有界变差函数.

下面我们要证明，$[a,b]$上的单调函数必在$[a,b]$上几乎处处有导数，所以有界变差函数也几乎处处有导数.这个重要结论证明较长，我们放在后面.现在先承认它，于是就可以回答第一个问题.

**定理 3(勒贝格定理)**　若 $f(x)$在$[a,b]$上勒贝格可积，则其不定积分 $F(x)=\int_a^x f(t)\mathrm{d}t$ 必为$[a,b]$上之有界变差函数，且几乎处处有

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\mathrm{d}t = f(x). \tag{7}$$

**证**　仿照定义勒贝格可积函数那样，把 $f(x)$写为

$$f(x)=f_+(x)-f_-(x),\ f_\pm(x)\geqslant 0,$$

则有

$$F(x)=\int_a^x f_+(t)\mathrm{d}t-\int_a^x f_-(t)\mathrm{d}t=F_+(x)-F_-(x).$$

因为 $f_\pm(x)\geqslant 0$，所以 $F_\pm(x)$均为单调不减.所以由定理 2，$F(x)$是有界变差函数，而且易见 $F_\pm(x)$都是连续函数.余下的仅是证明(7)式.这里我们需要一个引理.

**引理 4**　若 $f(x)$在$[a,b]$上为勒贝格可积，且对一切 $x\in[a,b]$，$\int_a^x f(t)\mathrm{d}t=0$，则在$[a,b]$上，$f(x)$几乎处处为 0.

**证**　用反证法，设结论不成立，从而可以找到一个正测度集 $E$，使 $f(x)>0$(或$<0$)于 $E$ 上，因为 $E^c$ 必可含于$[a,b]$的一个测度小于 $b-a$ 的开子集内，从而 $E$ 中必含一个具有正测度的闭子集，我们不妨设 $E$ 就是这个闭集，但由$\int_a^x f(t)\mathrm{d}t=0$对一切 $x\in[a,b]$成立，从而 $f(x)$ 在任一闭集 $E$ 上之积分为 0，因为这个积分将化为$[a,b]$上之积分(由假设为 0)以及一个开集 $E^c$ 上之积分的差.但 $E^c=\bigcup(\alpha_k,\beta_k)$，而$\int_{\alpha_k}^{\beta_k} f(t)\mathrm{d}t=\int_a^{\beta_k} f(t)\mathrm{d}t-\int_a^{\alpha_k} f(t)\mathrm{d}t=0$，从而$\int_{E^c} f(t$

所以用勒贝格控制收敛定理有

$$\lim_{l\to 0}\int_a^x \frac{F(t+l)-F(t)}{l}\mathrm{d}t=\int_a^x F'(t)\mathrm{d}t. \tag{9}$$

然而其左方的积分又等于$\frac{1}{l}\int_{a+l}^{x+l}F(t)\mathrm{d}t-\frac{1}{l}\int_a^x F(t)\mathrm{d}t=\frac{1}{l}\left[\int_x^{x+l}-\int_a^{a+l}\right]F(t)\mathrm{d}t$. 由 $F(x)$之连续性,可以把这两个积分均理解为黎曼积分,而知其极限是 $F(x)-F(a)=\int_a^x f(t)\mathrm{d}t$,代入(9)即得

$$\int_a^x [F'(t)-f(t)]\mathrm{d}t=0.$$

由引理 4 即知,当 $f(x)$为有界可积时,$F'(x)=f(x)$几乎处处成立.

为了对一般的可积(不一定有界)的 $f(x)$证明我们的结果,又需要另一个引理.

**引理 5** 若 $\varphi(x)$在$[a,b]$上连续非减,则 $\varphi'(x)$为勒贝格可积,且

$$\int_\alpha^\beta \varphi'(x)\mathrm{d}x\leqslant\varphi(\beta)-\varphi(\alpha),\ [\alpha,\beta]\subset[a,b]. \tag{10}$$

**证** 因为 $\varphi(x)$非减从而是有界变差函数,故 $\varphi'(x)$几乎处处存在. 但是另一方面 $\frac{\varphi(x+l)-\varphi(x)}{l}\geqslant 0$,因为 $\varphi(x)$是非减的. 故由法图引理

$$\lim_{l\to 0}\int_\alpha^\beta \frac{\varphi(x+l)-\varphi(x)}{l}\mathrm{d}x\geqslant\int_\alpha^\beta \lim\frac{\varphi(x+l)-\varphi(x)}{l}\mathrm{d}x=\int_\alpha^\beta \varphi'(x)\mathrm{d}x.$$

再由 $\varphi(x)$之连续性,仿照上面的证明又知左方为 $\varphi(\beta)-\varphi(\alpha)$,故引理证毕.

**定理 3 之证明的完成** 不妨设 $f(x)\geqslant 0$(即只看 $f_+(x)$). 作$[f(x)]_N$,有 $f(x)-[f(x)]_N\geqslant 0$,从而$\int_a^x\{f(t)-[f(t)]_N\}\mathrm{d}t$ 不减,连续且几乎处处有导数存在,其导数自然非负. 因此

$$\frac{\mathrm{d}}{\mathrm{d}x}\left\{\int_a^x f(t)\mathrm{d}t\right\}\geqslant\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x [f(t)]_N\mathrm{d}t. \tag{11}$$

但$[f(t)]_N$ 是有界可积的,故右方几乎处处为$[f(x)]_N$,左方则由 $F(x)$之定义为$\frac{\mathrm{d}}{\mathrm{d}x}F(x)$,于是

$F'(x)\geqslant[f(x$

在，于是我们考虑四个数

$$D^{+} f(x)=\overline{\lim_{l \to 0+}} \frac{1}{l}[f(x+l)-f(x)];$$

$$D_{+} f(x)=\underline{\lim_{l \to 0+}} \frac{1}{l}[f(x+l)-f(x)];$$

$$D^{-} f(x)=\overline{\lim_{l \to 0-}} \frac{1}{l}[f(x+l)-f(x)];$$

$$D_{-} f(x)=\overline{\lim_{l \to 0-}} \frac{1}{l}[f(x+l)-f(x)],$$

± 号分别表示右、左极限，写在 D 的上、下角分别表示上、下极限. 所以这四个数分别称为右上、……、左下导数. 所谓 $f(x)$ 在某点 $x$ 有导数，就是指在 $x$ 点的这四个数相等. 而只要有任何两个不相等，则 $f(x)$ 在 $x$ 点不可导. 然后，我们就分别考虑四个集合，例如

$$E=\{x: D_{+} f(x)<D^{-} f(x)\}$$

并证明 $m(E)=0$，这样我们将证得以下的极重要的定理：

**定理 6(勒贝格)**　若定义在 $[a, b]$ 上的函数 $f(x)$ 是单调的，则 $f'(x)$ 几乎处处存在.

这个定理虽然陈述十分明确，其证明则相当复杂. 下面我们只对 $f(x)$ 单调不减的情况证明，先证一个引理.

**引理 7(里斯)**　设 $G(x)$ 在 $[a, b]$ 上连续，如果在 $x \in [a, b]$ 右方有一点 $\xi$ 在 $[a, b]$ 中，使 $G(x)<G(\xi)$，则称 $x$ 为上升点，所有的上升点成为一个开集，而由许多区间组成，若 $(a_k, b_k)$ 是一个组成区间，则必有

$$G(a_k) \leqslant G(b_k). \tag{12}$$

**证**　因为 $G(x)$ 是连续的，故上升点 $x$ 的某个邻域中之点也都是上升点. 因此，上升点之集合相对于 $[a, b]$ 是开集，而由若干个开区间 $(a_k, b_k)$ 组成(但可能有一个是 $[a, b_k)$ 这也是一个相对开区间)，今证对一切 $x \in (a_k, b_k)$ 均有 $G(x) \leqslant G(b_k)$. 然后，令 $x \to a_k$ 即得(12)式. 在证明之前先提醒一下，在 $[b_k, b]$ 中的一切 $x$，均有

$$G(x) \leqslant G(b_k), \tag{13}$$

否则，$b_k$ 也是一个上升点，从而 $[b_k, b_k+\eta)$ ($\eta>0$ 是一个常数) 都是上升点，而 $(a_k, b_k)$ 不再是上升点集合的组成区间，$(a_k, b_k+\eta)$ 才是.

现证在 $(a_k, b_k)$ 内，(13) 也成立，我们用反

也是上升点，而如前所说，这是不可能的. 总之在 $[x_0', b]$ 中 $G(x) \leqslant G(b_k) < G(x_0')$，这与 $x_0'$ 是上升点相矛盾，从而引理得证.

现在开始证明定理 6，我们又分成几个步骤. 每一步分别考虑 $f'(x)$ 不存在的一种可能情况，并且证明每种情况都只能在一个零测度集上成立. 但是我们暂且仅限于 $f(x)$ 不仅单调而且连续的情况.

1. $E=\{x: \mathrm{D}^+ f(x)=+\infty\}$ 是零测度集之证明. 对 $E$ 中的 $x$，对任意固定的大数 $C$，必可找到 $\xi > x$ 使 $\dfrac{f(\xi)-f(x)}{\xi - x} > C$，亦即

$$f(x) - Cx < f(\xi) - C\xi.$$

这就是说，$E$ 中的 $x$ 必定都是 $f(x)-Cx$ 的上升点，因而 $E$ 是一个开集，而 $E$ 是若干开区间 $(a_k, b_k)$ 之并：$E = \bigcup (a_k, b_k)$，而由里斯引理

$$f(a_k) - Ca_k \leqslant f(b_k) - Cb_k, \text{即 } C(b_k - a_k) \leqslant f(b_k) - f(a_k).$$

对 $k$ 相加，注意到 $f$ 是单调不减的，故有

$$Cm(E) = C\sum_k (b_k - a_k) \leqslant \sum_k [f(b_k) - f(a_k)] \leqslant f(b) - f(a),$$

由 $C$ 之任意性即得 $m(E)=0$.

2. $E=\{x: \mathrm{D}^+ f(x) > \mathrm{D}_- f(x)\}$ 为零测度集之证明. 可以取 $c, C$ 为适合 $\mathrm{D}_- f(x) < c < C < \mathrm{D}^+ f(x)$ 之一切可能的有理数，而把 $E$ 分成可数多个集 $E_{c,C}$ 之并：

$$E_{c,C} = \{x: \mathrm{D}_- f(x) < c < C < \mathrm{D}^+ f(x)\},$$

再来证明每一个 $E_{c,C}$ 均为零测度集.

这时我们要把上升点的定义稍加修改. 其实前面讲的 $\xi > x$ 在 $x$ 右方，所以确切一点应该称这种 $x$ 为右上升点. 与此相应，若在 $x$ 左方某点 $\xi$ 处，即 $\xi < x$ 处，有 $G(\xi) > G(x)$，就称 $x$ 点为 $G(x)$ 的左上升点，而对于结论 $G(a_k) \geqslant G(b_k)$ 里斯引理仍成立. 于是对于 $x \in E$，由 $D_- f(x) < c$ 知必有 $\xi < x$ 使

$$\frac{f(\xi) - f(x)}{\xi - x} < c,$$

亦即

$$f(\xi) - c\xi > f(x) - cx \qquad (\text{注意 } \xi - x < 0),$$

而知 $x$ 是 $G(x) = f(x) - cx$ 之左上升点，仿照引理 7 的证明知 $E_{c,C} \subset \bigcup (a_k, b_k)$，而 $G(a_k) \geqslant G(b_x)$. 即

$$f(b_k) - f(a_k) \leqslant c(b_k - a_k). \tag{15}$$

现在我们限制取一个 $(a_k, b_k)$ 来研究，以它代替 $(a, b)$. 注意一个函数 $G(x)$ 在 $(a, b)$ 的右上升点之概念与它在 $(a_k, b_k) \subset (a, b)$ 中的右上升点概念不同. 对于前者，一定存在

$>C$，故对 $x\in E_{c,C}$ 必有 $\xi>x$ 使

$$\frac{f(\xi)-f(x)}{\xi-x}>C.$$

所以 $x$ 必是 $f(x)-Cx$ 相对于 $(a_k,b_k)$ 的右上升点，这样

$$(a_k,b_k)=\bigcup_j (a_{kj},b_{kj}),$$

而由里斯引理有

$$f(b_{kj})-f(a_{kj})\geqslant C(b_{kj}-a_{kj}).$$

对 $j$ 求和有

$$\sum_j (b_{kj}-a_{kj})\leqslant \frac{1}{C}\sum_j [f(b_{kj})-f(a_{kj})]\leqslant \frac{1}{C}[f(b_k)-f(a_k)].$$

再用(15)式并且对 $k$ 求和，就有

$$\sum_k\sum_j (b_{kj}-a_{kj})\leqslant \frac{c}{C}\sum_k (b_k-a_k)\leqslant \frac{c}{C}(b-a). \tag{16}$$

总之我们看到，对 $D_- f<c$ 和 $D^+ f>C$ 各用一次，即可把 $E_{c,C}$ 放在一个很细的覆盖 $\{(a_{kj},b_{kj})\}$ 中，且(15)成立.

再取一个 $(a_{kj},b_{kj})$，并在其中考查 $E_{c,C}$，又可以得到更精细的覆盖(现在小区间下有 4 个脚标了，因为 $D_- f<c$ 与 $D_+ f>C$ 又都各用了一次)：

$$\sum_k\sum_j\sum_l\sum_m (b_{kjlm}-a_{kjlm})\leqslant \frac{c}{C}\sum_k\sum_j (a_{kj}-a_{kj})\leqslant \left(\frac{c}{C}\right)^2 (b-a).$$

如此进行 $p$ 个循环会发现

$$m_e(E_{c,C})\leqslant \left(\frac{c}{C}\right)^p (b-a)\to 0. \tag{17}$$

所以 $E_{c,C}$ 是零测度集. 再对 $c,C$ 求和，因为一共只有可数多个 $E_{c,C}$，所以有 $m(E)=0$.

3. $E=\{x: D^- f(x)>D_+ f(x)\}$ 为零测度集之证明. 现在在与 $[a,b]$ 对称的区间 $[-b,-a]$ 上定义函数 $f^*(y)$ 如下(图 4-3-1)：

$$f^*(y)=-f(x),\ y=-x.$$

显然 $f^*(y)$ 在 $[-b,-a]$ 上是上升的，而且由前面的证明知

$$m\{y: D^+ f^*(y)=+\infty\}=0,$$

$$m\{y: D^+ f^*(y)>D_- f^*(y)\}=0.$$

但是若取 $k$ 充分小使 $y+k$ 仍在 $[-b,-a]$ 中($y$ 为端点时需要作的修改是自明的)，则

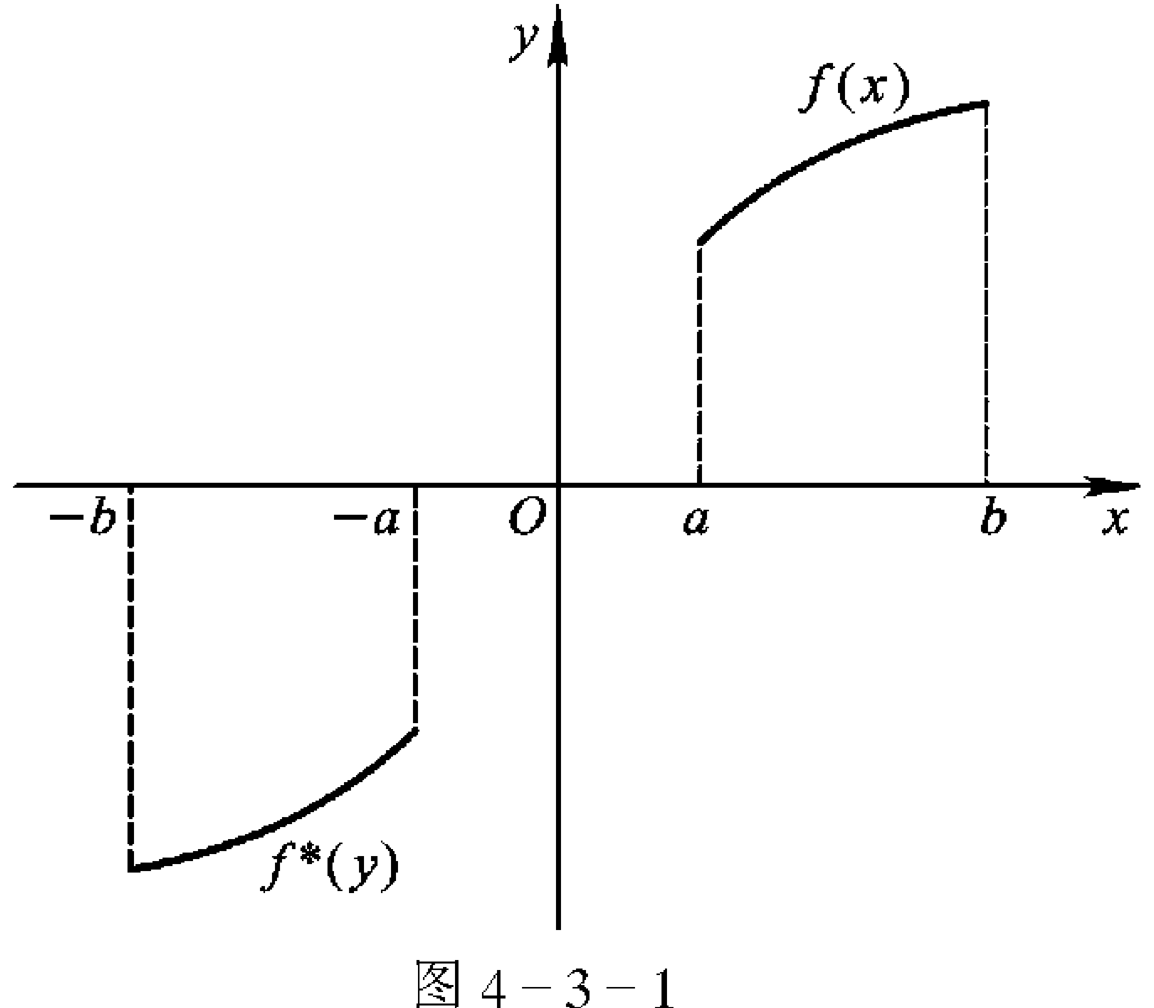



图 4-3-1

$$f^*(y+k)=-f(x-k),$$

$$\frac{f^*(y+k)-f^*(y)}{k}=-\frac{f(x-k)-f(x)}{k}.$$

记 $h=-k$，则当 $k\to 0\pm$ 时，$h\to 0\mp$，因此上式成为

$$\frac{f^*(y+k)-f^*(y)}{k}=\frac{f(x+h)-f(x)}{h}.$$

双方取上、下极限，即有

$$D^{\pm} f^*(y)=D^{\mp} f(x).$$

从而

$$m(E)=m\{x:D^- f(x)=+\infty\}=0$$

以及

$$m(E)=m\{x:D^- f(x)>D_+ f(x)\}=0. \tag{18}$$

4. 不连续情况的证明 总结前几段得知对于连续单调不减的 $f(x)$，1. 几乎处处有 $D^+ f(x)<+\infty$；2. 几乎处处有 $D^+ f(x)\leqslant D_- f(x)$；3. 几乎处处有 $D^- f(x)\leqslant D_+ f(x)$. 再加上自然就成立的下极限不大于上极限，我们终于得知

$$0\leqslant D^+ f(x)\leqslant D_- f(x)\leqslant D^- f(x)\leqslant D_+ f(x)\leqslant D^+ f(x)<+\infty$$

几乎处处成立. 最左方的"$0\leqslant$"来自 $f(x)$ 不减的假设，使得 $D^{\pm} f, D_{\pm} f\geqslant 0$. 由于这个不等式串首尾均是有限的 $D^+ f(x)$，所以在 $[a,b]$ 上，对于连续的不减函数 $f(x)$，几乎处处有

$$0\leqslant D_- f(x)=D_+ f(x)=D^- f(x)=D^+ f(x)<+\infty, \tag{19}$$

亦即这样的 $f(x)$ 在 $[a,b]$ 上几乎处处有导数 $f'(x)$. 于是对连续的不减函数 $f(x)$，勒贝格定理（定理 6）成立. 为了把这个结论推广到一般的非减函数，需要注意非减函数只有第一类间断点，若 $x_0$ 是一个间断点，则 $f(x_0-0)\leqslant f(x_0+0)$ 均存在，而 $f(x_0+0)-f(x_0-0)$ 是该点的跃度. 注意到 $\sum[f(x_0+0)-f(x_0-0)]\leqslant f(b)-f(a)$（$\sum$ 是对所有间断点求和），所以跃度 $\geqslant 1$ 的间断点为数有限；$\frac{1}{2}\leqslant$ 跃度 $<1$ 的间断点也只有有限个. 这样一来间断点最多有可数多个. 我们不要以为这可数多个间断点必将 $[a,b]$ 分成可数多个小区间，在每一个小区间中 $f(x)$ 均连续，因此可以利用上面的证明. 因为这些间断点可能是处处稠密的，所以它们把 $[a,b]$ 划分为可数多个小区间这个很直观的图像并不正确. §1 就举了一个这样的例子.（其实，上面讲的间断点可数的证明也不严格. 因为在不清楚它们是否可数之前，"$\sum$"的记号是不能用的，但读者不难自行改正.）所以，我们需要把里斯引理修改如下：设引理 7 中的 $G(x)$ 只是单调不减而不一定连续. 我们定义 $x\in[a,b]$ 为右上升点：如果在它的右方有点 $\xi>x$ 存在，且使

$$\max[G(x),G(x-0),G(x+0)]<G(\xi-0),$$

则称 $x$ 为右上升点，这种右上升点仍成开集，从而由可数多个开区间 $(a_k,b_k)$ 组成. 里斯引理指出

$$G(a_k+0)\leqslant G(b_k-0).$$

左上升点情况也一样，然后定理 6 的证明就不需改变了. 至此定理 6 证毕.

问题 Ⅰ 的解答至此已明白了. 下面转向问题 Ⅱ.

**3. 原函数，绝对连续性** 问题 Ⅱ 和问题 Ⅰ 的角度是不一样的. 问题 Ⅰ 是已知一个勒贝格可积的函数 $f(x)$，问它的积分 $\int_a^x f(t)\mathrm{d}t$ 有何性质. 答案是，它是一个在几乎处处意义下的原函数. 因此，只要任意找到一个绝对连续的（定义见后文）原函数 $F(x)$，必有

$$\int_a^b f(t)\mathrm{d}t = F(b) - F(a). \tag{20}$$

问题Ⅱ则是另一回事.已知 $F(x)$,问在何时 $F(x)$ 是一个勒贝格可积函数 $f(x)$ 的几乎处处意义下的原函数 $F'(x)=f(x)$ (a.e.),从而(20)成立?在问题Ⅰ的答案中已经看到,当 $f(x)$ 是勒贝格可积函数时,$F(x)=\int_a^x f(t)\mathrm{d}t$ 是连续的,而且具有界变差,所以自然会问,连续的有界变差函数是否问题Ⅱ的解答?很可惜,有反例指出并不如此.可以找到一个连续的不减函数 $F(x)$ 使得

$$\int_a^b F'(x)\mathrm{d}x < F(b) - F(a).$$

问题Ⅱ的解答是 $F(x)$ 是 $[a,b]$ 上的绝对连续函数,什么是绝对连续函数呢?

**定义 2** 设 $F(x)$ 定义在 $[a,b]$ 上,若对任一 $\varepsilon>0$ 均可找到 $\delta(\varepsilon)>0$ 使若 $[a,b]$ 的任意的有限多个并不重叠的区间 $[a_1,b_1],\cdots,[a_n,b_n]$ 之总长度小于 $\delta(\varepsilon)$,$F(x)$ 的下述变差小于 $\varepsilon$:

$$\sum_{k=1}^n |F(b_k) - F(a_k)| < \varepsilon, \tag{21}$$

就说 $F(x)$ 在 $[a,b]$ 上绝对连续.

绝对连续函数有一些简单的性质:

(1) 它是连续的,只要令 $n=1$ 就可以看到,因此它必然是有界的.

(2) 它具有有界变差,只要把 $[a,b]$ 分成若干小区间之并,使每一个小区间之长小于 $\delta$,然后在这些小区间上任作分划,即可用(21)式证明 $f(x)$ 在每个小区间上都有有界变差,从而在 $[a,b]$ 上也有有界变差.

(3) 绝对连续函数组成一个线性空间,两个绝对连续函数之积也是绝对连续的.

以上性质均容易证明,我们略去.

以下的目的是证明,问题Ⅱ的答案是:$F(x)$ 在 $[a,b]$ 上绝对连续.为了证明它,我们又需要一个看来几乎自明但实际上不太易证的引理.

**引理 8** 若 $F(x)$ 在 $[a,b]$ 上不减,且为绝对连续,则当 $F'(x)=0$ (a.e.)时,$F(x)\equiv$ const.

**证** 先说一下这个引理实际上不太易证明的原因.在通常的微积分教材中,由 $F'(x)=0$ 得证 $F(x)=\text{const}$ 是用了拉格朗日定理的,而这就要求 $F'(x)$ 处处存在而非几乎处处存在.$F(x)$ 的绝对连续性与单调性都只能保证 $F'(x)$ 几乎处处存在.

由于 $F(x)$ 不减,所以它映 $[a,b]$ 为区间 $\delta=[F(a),F(b)]$,今证 $m([F(a),F(b)])=0$,从而 $[F(a),F(b)]$ 缩为一点而引理得证.设 $Z$ 为使 $F'(x)$ 不存在或存在而不为 0 之点 $x$ 的集合,于是 $m(Z)=0$.记 $E$ 为 $Z$ 之余集,于是 $m(E)=b-a$,显然

$$\delta = F(Z)\cup F(E).$$

我们再分别证明 $F(Z)$ 与 $F(E)$ 均为 0 测度集.

对任一 $\varepsilon>0$,按绝对连续性条件找到 $\delta(\varepsilon)>0$,然后用一组(可能有可数多个)长度 $<\delta$ 的互不重叠(但端点可以相接)的小区间 $\{\Delta_k\}$ 把 $Z$ 覆盖起来,$Z\subset\Delta_1\cup\cdots\cup\Delta_n\cup\cdots$,显然 $F(Z)\subset\bigcup_{k=1}^\infty F(\Delta_k)$,但 $\bigcup_{k=1}^\infty F(\Delta_k)=\lim_{N\to\infty}\bigcup_{k=1}^N F(\Delta_k)$.由绝对连续性,$m\left(\bigcup_{k=1}^N F(\Delta_k)\right)<\varepsilon$.它的外极限

$\bigcup_{k=1}^{\infty} F(\Delta_k)$ 之外测度也小于 $\varepsilon$，$m_e(F(Z))<\varepsilon$ 自然也成立. 由 $\varepsilon$ 之任意性知 $F(Z)$ 有零测度.

至于 $F(E)$，则注意到对 $x\in E$，$F'(x)=0$，从而 $D^+F(x)<\varepsilon$，所以必有一个 $\xi>x$ 使

$$\frac{F(\xi)-F(x)}{\xi-x}<\varepsilon.$$

也就是说 $\varepsilon\xi-F(\xi)\geqslant\varepsilon x-F(x)$，而 $x$ 是 $\varepsilon x-F(x)$ 之右上升点. 由里斯引理(引理 6)，所有的右上升点之集合为若干个开区间 $(a_k,b_k)$ 之并，而且，$\varepsilon a_k-F(a_k)\leqslant\varepsilon b_k-F(b_k)$. 即有 $F(b_k)-F(a_k)<\varepsilon(b_k-a_k)$，对 $k$ 求和有

$$\sum_k[F(b_k)-F(a_k)]<\varepsilon\sum_k(b_k-a_k)\leqslant\varepsilon(b-a).$$

但 $F(E)\subset\bigcup_k[F(a_k),F(b_k)]$，所以，$m_e(F(E))<\varepsilon(b-a)$，而知 $F(E)$ 也是零测度集. 引理证毕.

这个引理有点怪：为什么要假设"$F(x)$ 不减"？似乎只要 $F'(x)=0$ 即应有 $F(x)\equiv\text{const}$. 关键是我们没有了拉格朗日公式. 在第三章我们就一再指出拉格朗日公式的重要性，这里又是一个证据. 看来，拉格朗日公式有点"娇嫩"，只要有一点不可求导，它就不适用了. 然而，最后我们还是可以抛弃 $F(x)$ 不减这个假设，由下面即将证明的定理 9，只要 $F(x)$ 绝对连续，而且 $F'(0)=0$ (a.e.)，则

$$F(b)-F(a)=\int_a^b F'(t)\mathrm{d}t=0.$$

**定理 9(勒贝格)** $F(x)$ a.e. 具有勒贝格可积的导数 $F'(x)$ 且使(20)成立(取 $f(x)=F'(x)$)之充分必要条件是，$F(x)$ 在 $[a,b]$ 上绝对连续.

证 必要性，对于任一 $\varepsilon>0$，取 $[a,b]$ 的有限多个互不重叠的子区间 $[a_k,b_k]$，记 $E=\bigcup_{k=1}^{N}[a_k,b_k]$，则

$$\sum_{k=1}^{N}\left|F(b_k)-F(a_k)\right|=\sum_{k=1}^{N}\left|\int_{a_k}^{b_k}F'(t)\mathrm{d}t\right|\leqslant\sum_{k=1}^{N}\int_{a_k}^{b_k}\left|F'(t)\right|\mathrm{d}t=\int_E\left|F'(t)\right|\mathrm{d}t.$$

但是由勒贝格可积函数的性质(iv)：积分对积分区域之连续性可知，一定存在 $\delta(\varepsilon)$，而当 $m(E)<\delta$ 时上式右方小于 $\varepsilon$，因此 $F(x)$ 绝对连续.

充分性. 设 $F(x)$ 为绝对连续，它必具有界变差，且因 $F(x)$ 连续，故可写为两个连续不减函数之差. 所以，不失一般性可设 $F(x)$ 为连续不减函数. 引理 5 告诉我们 $F'(x)$ 为可积的，而且

$$\int_\alpha^\beta F'(x)\mathrm{d}x\leqslant F(\beta)-F(\alpha).$$

由必要性的证明知 $G(x)=\int_a^x F'(t)\mathrm{d}t$ 也是绝对连续的，从而 $F(x)-G(x)$ 也是绝对连续的，它还是不减的，因此由引理 5

$$F(\beta)-F(\alpha)$$

所以由引理 8，$F(x)-G(x)=\text{const}$，而这个常数等于 $F(a)-G(a)$，故

$$F(x)=G(x)+F(a)-G(a)=F(a)+\int_a^x F'(t)\mathrm{d}t.$$

这就是(20)式，至此定理证毕.

经过这样复杂的讨论，可以把积分与微分互为逆运算这个命题的内涵归结如下：

Ⅰ 在黎曼连续函数框架下

$$\text{连续函数类} \underset{\frac{\mathrm{d}}{\mathrm{d}x}}{\overset{\int^x}{\rightleftarrows}} C^1\ \text{函数类}$$

Ⅱ 黎曼可积函数框架下

$$(R)\text{可积函数类} \underset{\frac{\mathrm{d}}{\mathrm{d}x}}{\overset{\int^x}{\rightleftarrows}} (?)$$

Ⅲ 勒贝格积分框架下

$$(L)\text{可积函数类} \underset{\frac{\mathrm{d}}{\mathrm{d}x}}{\overset{\int^x}{\rightleftarrows}} \text{绝对连续函数类}$$

但是在这两种积分框架下都得不到人们以为自然会成立的结论：若 $F'(x)=f(x)$，则 $F(x)=F(a)+\int_a^x f(t)\mathrm{d}t$. 即是说我们不会得到那么简单的结论：积分与微分互为逆运算.

在进入下一个问题之前，我们再讲一个十分有用的结果，即在勒贝格积分框架下的分部积分法.

**定理 10**(**分部积分法**) 设 $\Phi(x)$ 和 $\Psi(x)$ 是 $[a,b]$ 区间上的绝对连续函数，其导数是勒贝格可积函数 $\varphi(x)$ 与 $\psi(x)$，则有，

$$\int_a^b \Phi(x)\psi(x)\mathrm{d}x+\int_a^b \varphi(x)\Psi(x)\mathrm{d}x=\Phi(b)\Psi(b)-\Phi(a)\Psi(a). \tag{22}$$

**证** 由绝对连续函数的性质，$\Phi(x)\Psi(x)$ 也是绝对连续的，而且几乎处处有

$$\frac{\mathrm{d}}{\mathrm{d}x}[\Phi(x)\Psi(x)]=\Phi(x)\psi(x)+\varphi(x)\Psi(x),$$

双方在 $[a,b]$ 上积分即得(22)，证毕.

**注** 采用常见的写法，(22)就是

$$\int_a^b \Phi(x)\psi(x)\mathrm{d}x=\Phi(x)\Psi(x)\Big|_a^b-\int_a^b \varphi(x)\Psi(x)\mathrm{d}x. \tag{23}$$

**4. 二重积分与富比尼(Fubini)定理** 本书介绍测度只讲了一维空间中集合的测度. 原因是 $\mathbf{R}^n$ 中集合的测度从基本思想、基本性质来看虽与一维情况没有差别，计算要复杂得多. 其原因在于随维数的增加，集合的构造要复杂多了. 例如 $\mathbf{R}^1$ 中的开集必是至多可数多个开区间的并，但在高维空间则不然(这时的开区间要代以 $n$ 维开长方体 $\prod_{i=1}^n (a_i,b_i)$). $\mathbf{R}^n$ 中的勒贝格测度正是以 $n$ 维长方体的体积 $\prod_{i=1}^n (b_i-a_i)$ 为基础，而可以通过外测度等一连串手续像 $\mathbf{R}^1$ 中的测度一样

建立起来，而且也有可数可加性. 有了可测集以后就有可测函数理论，就有勒贝格积分理论，而前面的那些结果都是成立的. 正因为在这些基本方面都一致，为了避免过于冗长的计算，我们在前面就没有去讨论它，读者完全可以把 $\mathbf{R}^1$ 的理论平行地用于 $\mathbf{R}^n$. 现在也一样，我们只限于介绍 $\mathbf{R}^2$ 的情况，对于 $\mathbf{R}^n$ 的情况，读者自己会得出清晰的表述. 这里的重要结果的证明均略去.

$\mathbf{R}^n$ 中的测度与积分理论终究还有与 $\mathbf{R}^1$ 不同之处. 这就是化重积分为逐次积分的问题. 下面我们恒认为所要讨论的函数 $f(x,y)$ 是在 $\mathbf{R}^2$ 全空间或充分大的矩形 $A=[a,b]\times[c,d]$ 上求积分. 因为只要用积分区域 $\Omega$ 的特征函数 $\chi_\Omega(x,y)$ 去乘 $f(x,y)$，就可以把积分 $\iint\limits_{\Omega} f(x,y)\mathrm{d}x\mathrm{d}y$ 化为 $\iint\limits_{\mathbf{R}^2}\chi_\Omega f(x,y)\mathrm{d}x\mathrm{d}y$. 如果 $f(x,y)$ 是 $A$ 上的连续函数，则由通常的微积分教材知

$$\iint\limits_{A} f(x,y)\mathrm{d}x\mathrm{d}y=\int_a^b \mathrm{d}x\int_c^d f(x,y)\mathrm{d}y=\int_c^d \mathrm{d}y\int_a^b f(x,y)\mathrm{d}x. \tag{24}$$

但若 $f(x,y)$ 仅是黎曼可积，§1 定理(11)则只给出

$$\iint\limits_{A} f(x,y)\mathrm{d}x\mathrm{d}y=\int_a^b \mathrm{d}x\underline{\int_c^d} f(x,y)\mathrm{d}y=\int_a^b \mathrm{d}x\overline{\int_c^d} f(x,y)\mathrm{d}y$$

$$=\int_c^d \mathrm{d}y\underline{\int_a^b} f(x,y)\mathrm{d}x=\int_c^d \mathrm{d}y\overline{\int_a^b} f(x,y)\mathrm{d}x,$$

即内层积分换成了上、下积分. 现在问，如果把黎曼可积改为勒贝格可积又当如何？这时我们有

**定理 11(富比尼(Fubini)定理)** 设 $f(x,y)$ 在矩形 $A=[a,b]\times[c,d]$ 上勒贝格可积，则

(1) 若视 $f(x,y)$ 为 $y$ 之函数而 $x$ 为参数，则对几乎所有 $x\in[a,b]$，$f(x,y)$ 是 $y$ 在 $[c,d]$ 上的勒贝格可积函数，而且

$$I(x)=\int_c^d f(x,y)\mathrm{d}y$$

在 $[a,b]$ 上勒贝格可积.

(2) 将 $x$ 与 $y$ 对调，相应结果也成立.

(3) 二重积分可以化为逐次积分即有

$$\iint\limits_{A} f(x,y)\mathrm{d}x\mathrm{d}y=\int_a^b \mathrm{d}x\int_c^d f(x,y)\mathrm{d}y=\int_c^d \mathrm{d}y\int_a^b f(x,y)\mathrm{d}x.$$

把这个结果用于测度问题，我们知道 $\Omega\in A$ 为可测集的充分必要条件是特征函数 $\chi_\Omega(x,y)$ 为勒贝格可积，而且

$$m(\Omega)=\iint\limits_{A}\chi_\Omega(x,y)\mathrm{d}x\mathrm{d}y.$$

应用富比尼定理于 $\chi_\Omega(x,y)$ 有

$$m(\Omega)=\int_a^b \mathrm{d}x\int_c^d \chi_\Omega(x,y)\mathrm{d}y=\int_c^d \mathrm{d}y\int_a^b \chi_\Omega(x,y)\mathrm{d}x.$$

但是例如 $\int_c^d \chi_\Omega(x,y)\mathrm{d}y=m_\Omega(x)$ 表示 $\Omega$ 在直线 $x=$ 常数上截出的一维集合的线性测度，因此我们又有

**定理 12**　若 $\Omega\subset A$ 是可测集，则 $\Omega$ 与直线 $x=x_0$（或 $y=y_0$）之交线对几乎一切 $x_0$（或 $y_0$）为一维可测集，其测度 $m_\Omega(x_0)$ $(m_\Omega(y_0))$ 是 $x_0$（或 $y_0$）的勒贝格可积函数，而且

$$m(\Omega)=\int_a^b m_\Omega(x)\mathrm{d}x=\int_c^d m_\Omega(y)\mathrm{d}y. \tag{25}$$

(24)中两个逐次积分相等就表示当 $f(x,y)$ 为勒贝格可积函数时逐次积分可以变换次序．但如果没有定理中给出的条件，交换次序一般是不可能的．下面是一个例子：

$$\begin{aligned}\int_0^1\mathrm{d}x\int_0^\infty(\mathrm{e}^{-xy}-2\mathrm{e}^{-2xy})\mathrm{d}y&=\int_0^1\left[-\frac{1}{x}\mathrm{e}^{-xy}\Big|_0^\infty+\frac{1}{x}\mathrm{e}^{-2xy}\Big|_0^\infty\right]\mathrm{d}x\\&=\int_0^1\left(-\frac{1}{x}+\frac{1}{x}\right)\mathrm{d}x=0.\end{aligned}$$

但是

$$\int_0^\infty\mathrm{d}y\int_0^1(\mathrm{e}^{-xy}-2\mathrm{e}^{-2xy})\mathrm{d}x=\int_0^\infty\left[-\frac{1}{y}\mathrm{e}^{-xy}\Big|_0^1+\frac{1}{y}\mathrm{e}^{-2xy}\Big|_0^1\right]\mathrm{d}y$$
$$=\int_0^\infty\frac{\mathrm{e}^{-y}}{y}[\mathrm{e}^{-y}-1]\mathrm{d}y<0.$$

但是在一定条件下，逐次积分仍可交换积分次序：

**定理 13（托乃利(Tonelli)定理）**　若 $f(x,y)\geqslant0$ 为可测函数，则(24)式中的三个积分中只要有一个存在，另外两个一定也存在，且(24)式成立．

下一章中我们将再陈述这两个定理，但积分区域改为 $\mathbf{R}_x^n\times\mathbf{R}_y^n$．

最后我们还要提一下，多重积分的变量变换公式

$$\int_{\Omega_x}f(x)\mathrm{d}x=\int_{\Omega_y}f[x,y]\left|\frac{\partial(x^1,\cdots x^n)}{\partial(y^1,\cdots y^n)}\right|\mathrm{d}y$$

对勒贝格积分也成立，这里 $x=x(y)$ 是由 $\bar{\Omega}_y$ 到 $\bar{\Omega}_x$ 的 $C^1$ 映射且其逆也是．这个结论在最后一章中要反复用到，不过其证明涉及如何从测度的观点研究导数，于此只好暂略．

# §4　平方可积函数

**1. $L^2$ 空间的定义**　现在考虑一个问题：要测量某个物理量（例如气体压力）在不同条件（例如不同温度）下的值．设有 $N$ 个温度值，第一次测量的结果是

$$a_1^{(1)},a_2^{(2)},\cdots,a_N^{(1)},$$

第二次测量的结果是

$$a_1^{(2)},a_2^{(2)},\cdots,a_N^{(2)},$$

应该怎样评估两次测量的偏差？一个最直接的设想是计算每个温度下测量偏差之和：

$$\frac{1}{N}\sum_{k=1}^N(a_k^{(1)}-a_k^{(2)}).$$

其实就是平均值．但它显然不能准确反映实际情况：因为各项有正有负可能抵消．在极端的情况下，使得平均值可以为 0，而实际两次测量可以出现很大的误差．于是，好一点的方法是用

$$\frac{1}{N}\sum_{k=1}^{N}\ |a_k^{(1)}-a_k^{(2)}|.$$

它虽然避免了正负相消，但绝对值是很不便运算的.因此，更常见的是考虑

$$\Delta=\frac{1}{N}\sum_{k=1}^{N}\ [a_k^1-a_k^2]^2. \tag{1}$$

当然它的量纲不对，即物理量的单位不对，于是我们再开一次方得到

$$\delta=\left[\frac{1}{N}\sum_{k=1}^{N}(a_k^{(1)}-a_k^2)^2\right]^{\frac{1}{2}}. \tag{2}$$

$\delta$ 的量纲显然与测量值 $a_k^{(1)}, a_k^{(2)}$ 一致，因此它是合理的.

$\Delta$ 称为方差，$\delta$ 称为标准差.用方差和标准差来估量偏差，早在高斯以前就有了，而高斯系统地发展了它的理论特别是“最小二乘方”方法.可是我们要从另一个角度来看待它.

从数学上看采用(1)或(2)还有一个极大的好处，如果我们视一次测量的结果为 $N$ 维空间的点：$(a_1^{(1)},\cdots,a_N^{(2)})$ 和 $(a_1^{(1)},\cdots,a_N^{(2)})$，则(2)和(1)，除去因子 $\frac{1}{N}$ 后恰好是这两点的欧氏距离及其平方.这样，我们就可以用欧氏几何的概念和工具来处理它们.

以上考虑的是离散的情况，如果把观测压力对温度的变化连续进行，则会得到两个函数 $f^{(1)}(x)$ 和 $f^{(2)}(x)$，而(1)成为 $\int_a^b\ [f^{(1)}(x)-f^{(2)}(x)]^2\mathrm{d}x$.总之，我们会得到函数平方的积分.这种情况在物理学中是十分常见的.例如第三章变分法一节中我们看见电场的能也是一些函数的平方之和：$\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2$ 的积分.暂时不管导数问题(下一节我们还要仔细讨论)，则又是见到了函数平方的积分.总之，我们发现把一个函数平方再积分，用这个量来刻画函数性质，说明种种物理过程应该是十分有效的.

但是马上遇到的问题是：这里的积分是用黎曼积分好还是用勒贝格积分更好？从本章前面各节看，应该可以看得很清楚：应该用勒贝格积分.这一方面是因为我们在第一节中已经看到黎曼积分本质上说是适用于连续函数的积分，而物理学的发展，特别是因统计的思想引入了物理学，使得我们不能不把不连续函数或不光滑函数纳入我们的视野(回想一下布朗运动对数学家的启示)，而勒贝格积分理论正是适用于勒贝格可测函数的积分理论.这种测度和积分恰好适合概率论和统计学之需.另一方面则是因为勒贝格积分具有许多重要的性质，特别是完备性，所以本节将讨论以下的函数类.

**定义 1**　设 $f(x)$ 是 $E\subset\mathbf{R}$ 上的可测函数，而且 $|f(x)|^2$ 在 $E$ 上可积，这种函数之集合称为平方可积函数空间，记作 $L^2(E)$ (或简单记作 $L^2$).

**注 1**　是否要求 $E$ 为具有有限测度的可测集？有两种情况：或者要求 $E$ 是具有有限测度的可测集，例如 $E=\Omega\subset\mathbf{R}^n$ 是一有界开集；或者设有一串上升而穷竭 $\mathbf{R}^n$ 的子集序列 $\{\Omega_k\}$ 并使 $\mathbf{R}^n$ 的每个紧子集 $K$ 都含于某个 $\Omega_k$ 中.记 $E_k=E\cap\Omega_k$，于是 $E_1\subset E_2\subset\cdots$ 且 $E=\bigcup\limits_{k=1}^{\infty}E_k$ 是 $\{E_k\}$ 的外极限.每个 $E_k$ 均有有限测度.这时我们说 $E$ 具有 $\sigma$ 有限测度.在应用上它是很常见的，最常见的即 $E=\mathbf{R}^n$.当然，$\mathbf{R}^n$ 也是开集.

**注 2**　$L^2(E)$ 的函数既可取实值，也可取复值.在后一情况下 $|f(x)|^2=f(x)\cdot\overline{f(x)}$.

现在要问，若 $f(x)\in L^2(E)$，$f(x)$本身是否勒贝格可积？需知，定义中并未要求这一点.注意到

$$|f(x)|\leqslant\frac{1+|f(x)|^2}{2}$$

可见，若 $E$ 具有限测度，上式右方是可积的，因为$\frac{1}{2}$是可积函数.但若 $E$ 只是$\sigma$有限测度而本身并不具有限测度（我们也说 $m(E)=+\infty$），则上式不说明问题.以 $E=[1,\infty)\subset\mathbf{R}^1$ 为例，$f(x)=x^{-2/3}$是平方可积的，但本身并不可积.定义中一般不要求 $L^2(E)$函数本身为勒贝格可积原因在此.

关于 $L^2(E)$的性质，我们首先给出

**定理 1** $L^2(E)$是一个线性空间.

**证** 设 $f_1(x), f_2(x)\in L^2(E)$，$c_1, c_2$ 是实（或复）常数，则显然 $c_1f_1(x), c_2f_2(x)\in L^2(E)$，余下只需证明 $f_1(x)+f_2(x)\in L^2(E)$即可，这是很容易的，因为

$$|f_1(x)+f_2(x)|^2\leqslant(|f_1(x)|+|f_2(x)|)^2\leqslant 2|f_1(x)|^2+2|f_2(x)|^2.$$

而右方是可积的，左方至少是可测的，而且易证左方也是可积的.

**注** 若 $c$ 取复值，我们当然也应令 $f_1(x), f_2(x)$为复值（但 $x\in\mathbf{R}^2$ 仍为实的）函数.复值函数的积分很容易定义：只需其实，虚部均为可积即可，而且本章中一切结果（除专门规定适用于实值函数者——如正性——外）全部适用.量子物理的最基本的函数类（函数空间）就是复值的.

关于 $L^2(E)$函数之积则有

**定理 2** 若 $f(x), g(x)\in L^2(E)$，则 $f(x)g(x)$为勒贝格可积.

**证** 先看实 $L^2(E)$空间，因

$$(|f(x)|-|g(x)|)^2=|f(x)|^2+|g(x)|^2-2|f(x)g(x)|\geqslant 0,$$

故

$$|f(x)g(x)|\leqslant\frac{1}{2}|f(x)|^2+\frac{1}{2}|g(x)|^2.$$

左方是可测的，右方由定理 1 是两个可积函数之和故为可积的.因此结论成立.

这个证明对复 $L^2$ 空间自然仍成立，但我们习惯说 $f(x)\bar{g}(x)$可积.这当然只是一个记号问题，但是在量子力学中，其中却有深意.

我们在本节之始就提出，定义 $L^2$ 空间有一基本的思想，就是希望它尽可能与欧氏空间在几何上接近.一开始，我们就对离散的情况指出，一次测量的结果$(a_1^{(1)},\cdots,a_N^{(1)})$是 $\mathbf{R}^N$ 空间中一个向量，那么在连续的情况下一个 $L^2$ 函数 $f(x)$自然也可看作一个向量.（而且，现在 $x\in E$ 而$E$中一般有无穷多个点$x$，$f(x)$岂非应看作无穷维空间的向量（或点）吗？$f(x)$确实是一个无穷维空间中的向量，但是完全不是刚才讲的意义，这一点下面马上就要讲.）欧氏空间的几何学有两个基本的量，一是长度，二是角度.在离散情况下长度是 $\left(\sum_{k=1}^{N}\left|a_k^{(1)}\right|^2\right)^{1/2}$，那么现在是否应以 $\left(\int_E\left|f(x)\right|^2\mathrm{d}x\right)^{1/2}$ 作为其长度？至于两个向量的交角可以用内积 $\sum_{k=1}^{N}a_k^{(1)}\cdot a_k^{(2)}$ 来定义，现在是

否应定义两个 $L^2$ 函数 $f(x),g(x)$ 之内积为 $\int_E f(x)g(x)\mathrm{d}x$？确实如此.

**定义 2** 设 $f(x),g(x)\in L^2(E)$，我们定义 $f(x)$ 之模(或称范数，就是长度)为

$$\| f\|_{L^2} = \left(\int_E \left|f(x)\right|^2 \mathrm{d}x\right)^{1/2}. \tag{3}$$

定义实 $L^2$ 空间中 $f(x)$ 与 $g(x)$ 之内积为

$$\langle f,g\rangle_{L^2} = \int_E f(x)g(x)\mathrm{d}x. \tag{4}$$

复 $L^2$ 空间中的 $f(x)$ 与 $g(x)$ 之内积则定义为

$$(f,g)_{L^2} = \int_E f(x)\overline{g(x)}\mathrm{d}x. \tag{5}$$

**注** (4)和(5)都是对两个 $L^2$ 函数规定一个实或复数与之对应，都称为配对.(4)有时称为欧氏配对，用尖括号表示.(5)称为埃尔米特(C. Hermite——法国数学家)配对，用圆括号表示.但是有的书则记号相反.但不论是哪一种配对都有 $\langle f,f\rangle = \| f\|_{L^2}^2$，$(f,f) = \| f\|_{L^2}^2$，模或范数之下标在不发生误会时可以略去.

定义角度的根本公式是余弦定理，因为若有 $\mathbf{R}^N$ 空间的两个向量 $a,b$，则余弦定律是

$$a\cdot b = \| a\|\cdot\| b\|\cos\theta.$$

所以一定有 $\| a\|\cdot\| b\|\geqslant|a\cdot b|$.反之，只要有这样的不等式成立，则 $\dfrac{|a\cdot b|}{\| a\|\cdot\| b\|}\leqslant 1$，所以一定可找到一个角 $\theta$ 使 $a\cdot b = \| a\|\cdot\| b\|\cos\theta, 0\leqslant\theta\leqslant\pi$.所以，如果想把角度引入 $L^2(E)$，就应该找一个与余弦定理相应的不等式.

**定理 3**(**施瓦茨 Schwarz 不等式**) 若 $f(x),g(x)\in L^2(E)$ 为实值或是复值函数，则有以下的不等式成立

$$\int_E |f(x)g(x)|\mathrm{d}x\leqslant\left(\int_E\left|f(x)\right|^2\mathrm{d}x\cdot\int_E\left|g(x)\right|^2\mathrm{d}x\right)^{1/2}, \tag{6}$$

而且等号当且仅当 $|f(x)|$ 与 $|g(x)|$ 几乎处处相差一个常数因子时成立，

**证** 考虑 $\int_E(\lambda|f(x)|+|g(x)|)^2\mathrm{d}x$.这里 $\lambda$ 是实数.它当然是非负的，但是可将它写成 $\lambda$ 的二次三项式

$$A\lambda^2+2B\lambda+C\geqslant 0.$$

其中 $A=\int_E|f(x)|^2\mathrm{d}x, B=\int_E|f(x)|\cdot|g(x)|\mathrm{d}x, C=\int_E|g(x)|^2\mathrm{d}x$.因此必有 $B^2-AC\leqslant 0$，即(6)成立.若等号成立，即 $B^2-AC=0$，则当 $A\neq 0$ 时，必有一个 $\lambda$ 使 $A\lambda^2+B\lambda+C=0$，即

$$\int_E(\lambda|f(x)|+|g(x)|)^2\mathrm{d}x=0,$$

从而 $\lambda|f(x)|+|g(x)|=0$ 几乎处处成立，其逆亦真.若 $A=0$，则由 $B^2-AC=0$ 同时应有 $B=0$，这时应考虑 $|f(x)|+\mu|g(x)|\neq 0$，结论相同.

当然也就有

$$|\langle f,g\rangle_{L^2}|\ (\text{或}\ |(f,g)_{L^2}|)\leqslant\| f\|_{L^2}\ \| g\|_{L^2}. \tag{7}$$

既然 $f(x)\in L^2(E)$ 看作一个向量时 $\|f\|_{L^2}$ 表示其长，则 $\|f-g\|_{L^2}$ 应表示 $f$ 与另一个 $g\in L^2(E)$ 之距离：$\rho(f,g)$. 但是作为一个距离，应该满足以下三个条件：

(i)　$\rho(f,g)=\rho(g,f)$；

(ii)　$\rho(f,g)\geqslant 0$，$\rho(f,g)=0\Leftrightarrow f=g$；

(iii)（三角形不等式）$\rho(f,g)\leqslant\rho(f,h)+\rho(h,g)$，$h\in L^2(E)$.

用 $\|f-g\|$ 作为 $\rho(f,g)$ 时 (i) 自然是满足的，至于 (ii)，因为 $\|f-g\|^2=\int_E|f(x)-g(x)|^2\mathrm{d}x$，所以自然有 $\rho(f,g)\geqslant 0$. 由上一节讲的勒贝格积分的正性可知，$\rho(f,g)=0$ 时必有 $f(x)-g(x)=0$几乎处处成立. 但是在上一节中我们已经指出，两个几乎处处相等的勒贝格可积函数应看作相同的，所以 $f(x)-g(x)$ 与 0 应看作相同的. 在这个意义下 $f=g$.

最后证明三角形不等式

$$
\begin{aligned}
(\rho(f,g))^2 &= \int_E |f(x)-g(x)|^2\mathrm{d}x\leqslant\\
&= \int [(f-h)+(h-g)][\overline{(f-h)}+\overline{(h-g)}]\mathrm{d}x\\
&= \int |f-h|^2\mathrm{d}x+2\mathrm{Re}\int (f-h)\overline{(h-g)}\mathrm{d}x+\int_E |h-g|^2\mathrm{d}x\\
&\leqslant \|f-h\|^2+2\|f-h\|\cdot\|h-g\|+\|h-g\|^2\\
&= [\rho(f,h)+\rho(h,g)]^2.
\end{aligned}
$$

所以 (iii) 成立.

**2. $L^2$ 空间中的规范正交系**　至此，我们已看到 $L^2(E)$ 与欧氏空间非常相似. 欧氏空间中的正交坐标系是十分有用的工具. 就是说，在 $N$ 维欧氏空间 $E^N$ 中可找到 $N$ 个互相正交的单位向量 $e_1,\cdots,e_N$；即适合 $\langle e_i,e_j\rangle$（或 $(e_i,e_j)$）$=\delta_{ij}$. 因为它们是正交的，所以一定线性无关，证明如下：设有 $N$ 个常数 $c_1,\cdots,c_N$（如果考虑实欧氏空间，它们应为实数，如果考虑复欧氏空间——亦称埃尔米特空间 (Hermitian space)——则应取复数）使

$$c_1e_1+\cdots+c_Ne_N=0. \tag{8}$$

任取 $i$，$1\leqslant i\leqslant N$，并用 $e_i$ 与上式作欧氏的或埃尔米特的内积，则由 $\langle e_i,e_j\rangle=\delta_{ij}$（或 $(e_i,e_j)=\delta_{ij}$）有 $c_i=0$，$i=1,2,\cdots,N$. 因此 $e_1,\cdots,e_N$ 线性无关. 一个 $N$ 维线性空间 $\mathbf{R}^N$（不一定是欧氏空间，即我们不一定对它赋以长度与内积）的 $N$ 个线性无关元构成一个基底. 即是说 $\mathbf{R}^N$ 的任意元 $f$ 必可用它们线性表示. 实际上，设有任意元 $f\in\mathbf{R}^N$，则 $f,e_i,\cdots,e_N$ 是 $N+1$ 个元因此必线性相关：即有不全为 0 的常数 $c_0,c_1,\cdots,c_N$ 使

$$c_0f+c_1e_1+\cdots+c_Ne_N=0, \tag{9}$$

这里一定有 $c_0\neq 0$，否则会有不全为 0 的 $c_1,\cdots,c_N$ 使 (8) 成立，而这与 $e_1,\cdots,e_N$ 线性无关矛盾. 既然 (9) 中的 $c_0\neq 0$，则由 (9) 必有

$$f=\sum_{i=1}^{n}\alpha_ie_i,\quad \alpha_i=-\frac{c_i}{c_0}. \tag{10}$$

问题在于如何去求 $\alpha_i$？在欧氏空间（现在赋予了内积）中这个问题很容易解决. 事实上，$f$ 一定能用 $e_1,\cdots,e_N$ 表示，这是由 $\mathbf{R}^N$ 空间中最多只有 $N$ 个线性无关向量得出的，因此一定有 (10) 这样

的展开式，只不过其系数 $\alpha_1,\cdots,\alpha_N$ 待定而已. 为了定出这些系数，用 $e_i$ 与(10)之双方求内积，立即有

$$\alpha_i=\langle f,e_i\rangle,\quad 或\ \alpha_i=(f,e_i). \tag{11}$$

现在我们问，在 $L^2(E)$ 中能否建立与此平行的理论？首先我们要找到与上述 $e_i,\cdots,e_N$ 相应的对象.

**定义 3** 若两个 $L^2(E)$ 函数之内积(实或复)为 0，则说它们互相正交，若一个 $L^2(E)$ 函数之范数为 1 就称为归一的，规范的(unitary, normalized)(归一函数自然不会几乎处处为 0)，若一个 $L^2(E)$ 函数序列 $\{f_k(x)\}$ 适合 $\langle f_i,f_j\rangle=\delta_{ij}$ (或 $(f_i,f_j)=\delta_{ij}$)，则称 $\{f_k(x)\}$ 为一个归一或规范正交系(以下简记为 o. n. 系)

下面举几个例并设 $E\subset\mathbf{R}^1$. 最重要的 o. n. 系是 $L^2([-\pi,\pi])$ 上的 $\{\frac{1}{\sqrt{2\pi}}e^{ikx}\}$，$k=0,\pm1,\pm2,\cdots$，这是因为

$$\int_{-\pi}^{\pi}\frac{1}{\sqrt{2\pi}}e^{ikx}\overline{\frac{1}{\sqrt{2\pi}}e^{ilx}}dx=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{i(k-l)x}dx=\delta_{kl}. \tag{12}$$

由于它极端重要，我们将以第五章全章篇幅来讨论. 它其实就是通常微积分教材中讲的正、余弦函数的正交性，不过那里用的是 $\frac{1}{2\pi},\frac{1}{\sqrt{\pi}}\sin kx,\frac{1}{\sqrt{\pi}}\cos kx$，$k=1,2,\cdots$. 三角函数前的因子与 $\frac{1}{\sqrt{2\pi}}$ 不同是因为例如 $\cos kx=\frac{1}{2}(e^{ikx}+e^{-ikx})$，已经有了一个因子 2. 采用复数记号就没有了这个不整齐之处. 不过，引用复数更多地是因为量子力学的需要.

这种 o. n. 系在量子力学中起重要作用. $\left\{\frac{1}{\sqrt{2\pi}}e^{ikx}\right\}$ 是有限区间 $[-\pi,\pi]$ 上适用的，但量子力学中涉及无界区间时就不能用了，这时常用的是 $(-\infty,+\infty)$ 上的埃尔米特多项式：

$$H_n(x)=(-1)^n e^{x^2}\left(\frac{d}{dx}\right)^n e^{-x^2},\qquad n=0,1,2,\cdots. \tag{13}$$

它很明显是一个 $n$ 次多项式，为了证明它是一个正交系，我们先注意，它们恰好是 $e^{-s^2+2xs}$ 作为 $s$ 的函数(以 $x$ 为参数)在 $s=0$ 处的泰勒级数的系数(相差一个因子 $\frac{1}{n!}$)，这个函数称为埃尔米特多项式的生成函数. 事实上

$$e^{-s^2+2xs}=e^{x^2}e^{-(s-x)^2},$$

所以

$$\begin{aligned}
e^{-s^2+2xs}&=\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{d}{ds}\right)^n e^{-(s-x)^2}\cdot e^{x^2}\bigg|_{s=0}s^n\\
&=\sum_{n=0}^{\infty}\frac{e^{x^2}}{n!}\left(\frac{d}{ds}\right)^n e^{-(s-x)^2}\bigg|_{s=0}s^n\\
&=\sum_{n=0}^{\infty}\frac{e^{x^2}}{n!}\left(-\frac{d}{dx}\right)^n e^{-(s-x)^2}\bigg|_{s=0}s^n\\
&=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}e^{x^2}\left(\frac{d}{dx}\right)^n e^{-x^2}s^n=\sum_{n=0}^{\infty}\frac{H_n(x)}{n!}s^n.
\end{aligned}$$

为了研究$\{H_n(x)\}$之正交性，考虑$e^{-s^2+2xs}\cdot e^{-t^2+2xt}=\sum_{m,n=0}^{\infty}\frac{1}{m!\ n!}H_m(x)H_n(x)t^m s^n$. 双方乘以$e^{-x^2}$再对$x$在$(-\infty,+\infty)$上积分. 因为$H_m(x)H_n(x)$是$x$的$m+n$次多项式，乘上$e^{-x^2}$后在$x\to\pm\infty$时函数迅速趋于0，而上述级数对$x$一致收敛(也可以用勒贝格控制收敛定理，但是视$t,s$为参数)有

$$\int_{-\infty}^{\infty}e^{-x^2}e^{-s^2+2xs-t^2+2xt}dx$$
$$=\int_{-\infty}^{\infty}e^{2ts}e^{-(x-t-s)^2}dx=\sum_{m,n=0}^{\infty}\frac{t^m s^n}{m!\ n!}\int_{-\infty}^{\infty}e^{-x^2}H_m(x)H_n(x)dx.$$

利用著名的高斯积分$\int_{-\infty}^{\infty}e^{-\xi^2}d\xi=\sqrt{\pi}$(这个积分太重要了，我们下一节专门讨论它)，即知式左为

$$e^{2ts}\int_{-\infty}^{\infty}e^{-(x-t-s)^2}dx=e^{2ts}\int_{-\infty}^{\infty}e^{-\xi^2}d\xi\qquad(\xi=x-t-s)$$
$$=\sqrt{\pi}e^{2ts}=\sqrt{\pi}\sum_{n=0}^{\infty}\frac{2^n}{n!}t^n s^n.$$

与上式右方比较，注意到若两个幂级数相等则其同次幂系数也相等(对于两个自变量$t$和$s$的二重幂级数$\sum_{m,n=0}^{\infty}a_{mn}t^m s^n$也容易证明$a_{mn}$为其和$F(t,s)$的导数$\frac{\partial}{m!\ n!}\frac{\partial^{m+n}F(0,0)}{\partial t^n\partial s^n}$)，即有

$$\int_{-\infty}^{\infty}e^{-x^2}H_m(x)H_n(x)dx=\sqrt{\pi}2^n\cdot n!\ \delta_{mn}.\tag{14}$$

我们称此式为$\{H_m(x)\}$以$e^{-x^2}$为权的正交性. 引入了权函数$\rho(x)\geqslant 0$后，我们定义

$$\langle f,g\rangle=\int_E\rho(x)f(x)g(x)dx\tag{15}$$

和

$$\|f\|^2=\langle f,f\rangle=\int_E\rho(x)|f(x)|^2dx\tag{16}$$

为以$\rho(x)$为权的内积和范数($(f,g)$之定义相仿)，而且称使得(16)中的$\|f\|<+\infty$的可测函数集合为以$\rho$为权的$L^2(E)$空间，它与上面讲的$L^2(E)$性质相似.

但由(14)可知$\{H_m(x)\}$没有规范化，为此我们再添加一个规范化因子(归一化因子)$N_m=(\sqrt{\pi}2^n n!)^{-\frac{1}{2}}$，而称$\frac{(-1)^n}{(\sqrt{\pi}2^n\cdot n!)^{1/2}}e^{x^2}\left(\frac{d}{dx}\right)^n e^{-x^2}$为规范化(归一化)埃尔米特多项式.

$H_m(x)$是以下微分方程

$$u''-2xu'+2mu=0$$

的以$e^{-x^2}$为权的$L^2(\mathbf{R}^1)$解. 量子力学中常用的许多o.n.系都是这样来的. 例如当我们讨论具有球对称性的量子力学问题时会遇到要寻求

$$\frac{d}{dx}\left[(1-x^2)\frac{du}{dx}\right]+m(m+1)u=0\tag{17}$$

(勒让德方程)在$[-1,1]$上的多项式解，这个方程的特点是在区间两端$x=\pm 1$处有奇点.

可以证明

$$P_m(x)=\frac{1}{2^m\cdot m!}\frac{\mathrm{d}^m}{\mathrm{d}x^m}(x^2-1)^m \tag{18}$$

是方程(17)的解，也可证明$\{P_m(x)\}$是一个正交系：

$$\int_{-1}^{1}P_m(x)P_n(x)\mathrm{d}x=\frac{2}{2m+1}\delta_{mn},$$

因此，添加归一化因子$\sqrt{\dfrac{2m+1}{2}}$后，$\left\{\sqrt{\dfrac{2m+1}{2}}P_m(x)\right\}$就成了一个 o. n. 系. (18)称为勒让德多项式 $P_m(x)$的罗德里格斯(Rodrigues)公式.

举了以上几个例子以后，设已有 $L^2(E)$的 o. n. 系$\{f_n(x)\}$，任一个 $L^2(E)$函数 $f(x)$可否仿照 $\mathbf{R}^N$ 中的坐标展开式(10)写为

$$f(x)=\sum_{n=0}^{\infty}\alpha_n f_n(x)? \tag{19}$$

如果可能，又如何求系数 $\alpha_n$？至少在一个情况下可知(19)式是不可能的：如果在 $\mathbf{R}^N$ 中从坐标系$\{e_1,\cdots,e_N\}$中删去若干个而得例如$\{e_2,\cdots,e_N\}$，则(10)式是不可能，因为这时(10)式右方成为$\sum\limits_{i=2}^{N}\alpha_i e_i$ 是与$e_1$ 正交的，它们只构成一个 $N-1$ 维空间. 所以 $\sum\limits_{i=2}^{N}\alpha_i e_i$ 只能表示与$e_1$ 正交的$N-1$维子空间中的全体向量，而对 $\mathbf{R}^N$ 中的全部向量它不可能成立. 从几何上这是不言而喻的. 现在 $L^2(E)$在一定意义下是无穷维向量空间(其 o. n. 系含有无穷多个元，上面三个例子都是). 如果删除了 o. n. 系中的若干个元，使得它“不完全”了，则至多它只能表示与被删去的元正交的那些向量. 这个删减的 o. n. 系称为“不完全的”. 在有限维空间，一个 o. n. 系是否完全很容易判断：完全性的充要条件是系中向量个数等于空间维数，但在无限维情况，删去若干向量后，余下的可能仍是无穷个. 因此我们要从另一个角度来讨论这个问题：设有 o. n. 系$\{f_n(x)\}$，任给一个$f(x)\in L^2(E)$，并且确定正整数 $N$，在长度为 $N$ 的线性组合 $\sum\limits_{k=1}^{N}c_k f_k(x)$中，怎样选系数$\{c_1,\cdots,c_N\}$能使$\left\|f-\sum\limits_{k=1}^{N}c_k f_k\right\|$最小？

回答这个问题其实很容易：(为简单计设 $f,f_k$ 均为实值函数)经过一些计算，可以得出

$$\begin{aligned}\|f-\sum_{k=1}^{n}c_k f_k\|^2&=\|f\|^2-2\sum_{k=1}^{N}(\alpha_k c_k)+\sum_{k=1}^{N}c_k^2\\&=\|f\|^2-\sum_{k=1}^{N}\alpha_k^2+\sum_{k=1}^{N}(\alpha_k-c_k)^2\geqslant\|f\|^2-\sum_{k=1}^{N}\alpha_k^2.\end{aligned} \tag{20}$$

这里 $\alpha_k=\langle f,f_k\rangle$，称为 $f$ 对$\{f_k\}$的傅里叶系数. 所以当且仅当线性组合 $\sum\limits_{k=1}^{N}c_k f_k$ 之系数为$f$ 之傅里叶系数时，$\|f-\sum\limits_{k=1}^{N}c_k f_k\|$将达到其最小值$\|f\|^2-\sum\limits_{k=1}^{N}\alpha_k^2$. 这里最有趣的是，若把 $N$ 增到大$M$，再求类似问题之解时，则前 $N$ 个系数$c_k=\alpha_k$ 不改变，而只要调整 $c_{N+1},\cdots,c_M$ 使之等于$\alpha_{N+1},\cdots,\alpha_M$ 即可，所以我们可以令 $N\to\infty$而有

**定理 4**　若 $f(x)\in L^2(E)$而$\{f_k(x)\}$为 $L^2(E)$中的 o. n. 系，则有以下的贝塞尔(Bessel)不

等式成立：

$$\|f\|^2\geqslant\sum_{k=1}^{\infty}\alpha_k^2. \tag{21}$$

**证** 在(20)中令 $c_k=\alpha_x$，因式左非负，故

$$\|f\|^2-\sum_{k=1}^{N}\alpha_k^2\geqslant 0.$$

令 $N\to\infty$，即知 $\sum_{k=1}^{\infty}\alpha_k^2<+\infty$ 且(21)成立.

**注** 如果考虑复的 o.n.系 $\{f_k\}$，则 $\alpha_k^2$ 应改为 $|\alpha_k|^2$.

注意到在 $\mathbf{R}^N$ 中，两个向量之内积 $a\cdot b$ 即 $a$ 在 $b$ 上之投影.在 $L^2(E)$ 中，如果我们视 o.n.系 $\{f_n(x)\}$ 是一组单位长的互相正交的向量，即正交坐标系，则 $f(x)\in L^2(E)$ 的傅里叶系数就是向量 $f$ 在这些单位坐标向量上的投影.在 $\mathbf{R}^N$ 中，按毕达哥拉斯定理，如果这些单位正交坐标向量没有少取的话，则投影的平方和应该就是是向量长平方.与此类比，若对任意 $f(x)\in L^2(E)$，(21)式中有等号成立，则应该标志着 o.n.系 $\{f_n(x)\}$ 中没有“缺少”了哪一个向量.

**定义 4** 若 $\{f_k(x)\}$ 是 $L^2(E)$ 中的 o.n.系，若再没有一个非 0(即不处处为 0)的 $L^2(E)$ 函数与一切 $f_k(x)$ 正交，就说 $\{f_k(x)\}$ 是完全的 o.n.系.

**定理 5** o.n.系 $\{f_k(x)\}$ 为完全的充分必要条件是(21)式对任意 $f(x)\in L^2(E)$ 有等号成立：

$$\|f\|^2=\sum_{k=1}^{\infty}|\alpha_k|^2,\qquad \alpha_k=(f,f_k). \tag{22}$$

(22)式称为封闭性方程或帕塞瓦尔(Parseval)等式.

**证** **充分性** 若(22)成立而有非 0 的 $f(x)\in L^2(E)$ 与一切 $f_k(x)$ 正交，因此 $\alpha_k=(f,f_k)=0$，而由(22)有 $f$ 几乎处处为 0，这与 $f$ 非 0 矛盾.

**必要性** 今设 o.n.系 $\{f_k(x)\}$ 是完全的，证明(22)成立.实际上，任取 $f(x)\in L^2(E)$，而 $\alpha_k=(f,f_k)$，我们来看级数 $\sum_{k=1}^{\infty}\alpha_k f_k$.首先的问题是这个级数在什么意义下收敛.为此，我们形式地记 $g(x)=\sum_{k=1}^{\infty}\alpha_k f_k(x)$，并取此级数的部分和序列 $S_N(x)=\sum_{k=1}^{N}\alpha_k f_k(x)$.易见 $\|S_N(x)-S_M(x)\|^2=\sum_{k=N+1}^{M}|\alpha_k|^2$(设 $M>N$).由贝塞尔不等式，$\sum_{k=1}^{\infty}|\alpha_k|^2<+\infty$，故 $N,M$ 充分大时 $\|S_N(x)-S_M(x)\|<\varepsilon$.由下面就要证明的里斯-费希尔(Riesz-Fisber)定理(关于勒贝格可积函数的里斯-费希尔定理即上节之理 19，它们的重要性在上节已讲到了.在 $L^2(E)$ 中这定理也成立，也同样极为重要，见下文)知 $S_N(x)$ 在 $L^2(E)$ 中收敛于一个 $L^2(E)$ 函数，记它为 $g(x)$，此即 $g(x)=\sum_{k=1}^{\infty}\alpha_k f_k(x)$ 的意义.现在再来计算 $g(x)$ 之傅里叶系数.因 $\|S_N(x)-g(x)\|\to 0$.取 $N>k$，有 $(g,f_k)=(g-S_N,f_k)+(S_N,f_k)=(g-S_N,f_k)+a_k$.令 $N\to\infty$，|

证明的$L^2(E)$中的里斯-费希尔定理，$\|S_N\|\to\|f\|$，但$\|S_N\|^2=\sum_{k=1}^{N}\alpha_k^2$，令$N\to\infty$即知
式成立.

给出了一个o.n.系，怎样判定它完全与否？这时常是最吃力的问题，因为应用定理5
方便，而且这常是最重要的问题.下一章我们就要证明$\{e^{ilx}\}$是$L^2([-\pi,\pi])$中的完全系.现
们则回到$L^2(E)$理论中一个最重要的定理.§2中我们已对勒贝格可积函数证明过一个同
定理(定理19).

**定理6(里斯－费希尔定理)**　设$\{f_k(x)\}$是$L^2(E)$中的一个柯西序列，即对任一$\varepsilon>0$
$N(\varepsilon)$存在，使当$n,m>N(\varepsilon)$时$\|f_n-f_m\|<\varepsilon$，则必存在唯一的$f(x)\in L^2(E)$，使$\|f_n-$
0，而且$\|f_n\|\to\|f\|$.

**证**　§2中证明定理19时，我们是假设了$m(E)<+\infty$的，而只在最后提了一句话当
$\sigma$有限时定理也成立，现在我们给出完全的证明.

(i)先设$m(E)<+\infty$，注意这时$L^2(E)$函数本身也为勒贝格可积.现在取$N_1<N_2<$
$N_k\cdots$，使

$$\|f_{N_{k+1}}-f_{N_k}\|<\frac{1}{2^k},$$

而由施瓦茨(Schwarz)不等式

$$\int_E|f_{N_{k+1}}-f_{N_k}|\mathrm{d}x\leqslant|m(E)|^{1/2}\|f_{N_{k+1}}-f_{N_k}\|_{L^2}.$$

所以$\sum_{k=1}^{\infty}\int_E|f_{N_{k+1}}-f_{N_k}|\mathrm{d}x<+\infty$，而依照上一节勒维定理的证明可知，$f_{N_1}($
$\sum_{k=1}^{\infty}[f_{N_{k+1}}(x)-f_{N_k}(x)]$几乎处处收敛于某个函数$f(x)$，亦即$\{f_k(x)\}$有一个子序列$\{f_{N_k}$
在$E$上几乎处处收敛于$f(x)$.$f(x)$在$E$上显然是可测的，今证它属于$L^2(E)$.为此，也
证明定理19一样，可知当$n>N(\varepsilon)$时

子集 $E_1$ 上几乎处处收敛于 $f(x)$，所以 $\underset{\sim}{f}(x)=f(x)$ 于 $E_1$ 上，或者说 $f(x)$ 可以延拓到 $E_2$ 上仍为可测函数. 我们仍记它为 $f(x)$（上面的记法是 $\underset{\sim}{f}(x)$）. 用对角线法即知在 $E$ 上有一可测函数 $f(x)$，使在每个 $E_k$ 上都可以仿照上面的证明而有

$$\int_{E_k}\left|f_n(x)-f(x)\right|^2\mathrm{d}x<\varepsilon^2. \tag{24}$$

暂时固定 $n$，由 $f(x)=f_n(x)+f(x)-f_n(x)$，用三角形不等式有

$$\begin{aligned}\left(\int_{E_k}\left|f(x)\right|^2\mathrm{d}x\right)^{\frac{1}{2}}&\leqslant\varepsilon+\left(\int_{E_k}\left|f_n(x)\right|^2\mathrm{d}x\right)^{\frac{1}{2}}\\&\leqslant\varepsilon+\left(\int_{E}\left|f_n(x)\right|^2\mathrm{d}x\right)^{\frac{1}{2}}.\end{aligned}$$

再令 $k\to\infty$，即知 $f(x)\in L^2(E)$. 在(24)中先令 $k\to\infty$，再令 $n\to\infty$ 即知

$$\lim_{n\to\infty}\|f_n-f\|=0,$$

即是说 $L^2$ 收敛也成立.

定理的最后一个论断并非此定理之一部分只是顺便放在这里，因为证明十分容易：利用

$$f_n(x)=f_n(x)-f(x)+f(x);\ f(x)=f(x)-f_n(x)+f_n(x).$$

由三角形不等式有

$$\|f_n\|\leqslant\|f\|+\|f_n-f\|,\quad\|f\|\leqslant\|f_n\|+\|f_n-f\|,$$

亦即

$$\big|\,\|f_n\|-\|f\|\,\big|\leqslant\|f_n-f\|\to0.$$

至此我们已经完成了对 $L^2(E)$ 的讨论. 可是在历史上，首先讨论的是另一个空间 $l^2$，这是希尔伯特与施密特在研究具有对称核的积分方程理论时提出来的. $l^2$ 是一个由复数序列组成的空间. 在 $L^2(E)$ 中可以找到完全的 o.n. 系（这一点证明尚未完成）$\{f_k(x)\}$ 后，任意 $f(x)\in L^2(E)$ 就都可以展开为以下形状的级数

$$f(x)=\sum_{k=1}^{\infty}\alpha_kf_k(x),\ \alpha_k=\langle f,f_k\rangle\ (\text{或 }\alpha_k=(f,f_k)). \tag{25}$$

我们不妨称它为 $f(x)$ 之关于 o.n. 系 $\{f_n(x)\}$ 的傅里叶级数. 上面证明了它的部分和在(23)式意义下收敛于 $f(x)$，即此级数在 $L^2$ 意义下收敛于 $f(x)$. 帕塞瓦尔等式指出

$$\|f\|=\left(\sum_{k=1}^{\infty}\left|\alpha_k\right|^2\right)^{1/2}. \tag{26}$$

(22)式是对实的 $L^2(E)$ 系数讲的，不过这个理论对复值 $L^2$ 函数也成立，这时帕塞瓦尔等式应改写如上，于是我们看见 $L^2(E)$ 函数 $f(x)$ 与一个序列对应：

$$f(x)\sim(\alpha_1,\alpha_2,\cdots,\alpha_k,\cdots).$$

很容易看到，这个对应是双方 $1-1$ 的，而且保持线性空间结构，如同线性代数中讲的线性同构一样. 式左的 $f(x)\in L^2(E)$，故有内积与范数，如若还有另一个 $g(x)\in L^2(E)$ 对应于 $(\beta_1,\beta_2,\cdots,\beta_k,\cdots)$，则由 $f(x)=\sum_{k=1}^{\infty}\alpha_kf_k(x)$，$g(x)=\sum_{k=1}^{\infty}\beta_kf_k(x)$，可以证明

$$\langle f,g\rangle=\sum_{k=1}^{\infty}\alpha_k\beta_k,\ \|f\|^2=\sum_{k=1}^{\infty}\alpha_k^2,\ \|g\|^2=\sum_{k=1}^{\infty}\beta_k^2.$$

因此,我们在集合$\{\alpha_1,\alpha_2,\cdots,\alpha_k,\cdots\}$中也可以引入内积(实的和复的)和范数

$$\langle\alpha,\beta\rangle=\sum_{k=1}^{\infty}\alpha_k\beta_k\ (\text{或}\ (\alpha,\beta)=\sum_{k=1}^{\infty}\alpha_k\bar{\beta}_k),$$

$$\|\alpha\|^2=\sum_{k=1}^{\infty}\alpha_k^2,\ (\text{或}\ \|\alpha\|^2=\sum_{k=1}^{\infty}|\alpha_k|^2),\tag{27}$$

这里 $\alpha=(\alpha_1,\alpha_i,\cdots,\alpha_k,\cdots)$,$\beta=(\beta_1,\beta_2,\cdots,\beta_k,\cdots)$. 在这个集合中引入线性结构和由上式定义的内积与范数后,我们得到一个空间,记为 $l^2$. $l^2$ 与 $L^2(E)$线性同构. 这个同构由 $L^2(E)$中的一个完全的 o.n. 系来实现,选用不同的完全 o.n. 系会得到不同的同构,但是总是同构. 这个同构又是等距的,这一点由(26)式以及很容易证明的

$$\langle f,g\rangle=\sum_{k=1}^{\infty}\alpha_k\beta_k,\ (f,g)=\sum_{k=1}^{\infty}\alpha_k\bar{\beta}_k$$

来实现. 不问那个完全的 o.n. 系如何选择,上面的关系式总是不变的. 因为 $L^2(E)$与 $l^2$ 有上述的等距同构关系,我们把二者看作完全相同的空间. 而且上面已提到,在历史上是先有 $l^2$,后来才由里斯提出 $L^2(E)$. $l^2$ 中也有里斯–费希尔定理,但我们不来证明它了. 读了第六章中关于完备性的论述就会明白.

至此也可以回答一个问题:$L^2(E)$是无穷维空间应如何理解? 这不是说 $f(x)\in L^2(E)$而 $E$ 中的 $x$ 一般有无穷多个点就说 $f(x)$是无穷维向量,而是因为 $L^2(E)$与 $l^2$ 同构,而后者显然有一个含无穷多个线性无关向量的基底,例如$\{(1,0,\cdots,0,\cdots);(0,1,0,\cdots,0,\cdots)\cdots;(0,\cdots,0,1,0,\cdots)\cdots\}$,才说 $l^2$ 是无穷维向量. 所以与它同构的 $L^2(E)$也有无穷维;而上述 $l^2$ 的基底(或坐标)就对应于 $L^2(E)$的一个完全 o.n. 系$\{f_1(x),f_2(x),\cdots,f_k(x),\cdots\}$.

$L^2(E)$空间具有无穷维的欧氏空间(或埃尔米特空间)的结构,因此是性质极丰富而简单性又仅次于 $\mathbf{R}^N$ 的空间,它是量子力学的基本框架.

**3. $L^p$ 空间** 研究了 $L^2(E)$后,就可以考虑 $L^p(E)$和 $l^p$,这里 $1\leqslant p\leqslant\infty$. $L^P(E)$就是 $p$ 次幂勒贝格可积的函数之空间. 这时会问当 $p>1$ 时,若$|f(x)|^p$ 勒贝格可积,$f(x)$本身又如何? 和 $L^2(E)$一样,若 $m(E)<+\infty$,则 $f(x)$本身也勒贝格可积,否则就不一定了. 当然 $p=1$ 时,$L^p(E)=L^1(E)$即勒贝格可积函数空间,而且通常也喜欢用这个记号而不用 $L(E)$. $L^\infty(E)$则是另一回事. $L^p(E)$也是线性空间,这一点不难证,我们暂时略去.

下面限于 $1<p\leqslant\infty$,这个限制极为重要.

$L^p$,$p\neq2$ 和 $L^2$ 最大的差别在于 $L^p(p\neq2)$中没有内积,因此就没有 o.n. 系这一重要概念. 但在 $L^p$ 中却有与施瓦茨不等式相应的重要不等式. 以下我们说,适合关系式$\dfrac{1}{p}+\dfrac{1}{q}=1$ 的 $p$ 和 $q$(这里 $1<p,q<\infty$)为共轭指数,而 $L^p$ 与 $L^q$ 也称为互相共轭的空间. 当然,$p=1$ 时,$q=\infty$,所以 $L^1$ 和 $L^\infty$ 也互相共轭,但其性质与 $1<p,q<\infty$ 极不相同. 显然,当 $p=2$ 时也有 $q=2$,所以 $L^2$ 是自共轭的.

**定理 7 (赫尔德(Hölder)不等式)** 设 $f(x)\in L^p(E)$,$g(x)\in L^q(E)$,$\dfrac{1}{p}+\dfrac{1}{q}=1$,$1<p,\ q<\infty$,则 $f(x)g(x)\in L^1(E)$,而且

$$\int_E |f(x)g(x)|\mathrm{d}x \leqslant \left(\int_E |f(x)|^p \mathrm{d}x\right)^{1/p}\left(\int_E |g(x)|^q \mathrm{d}x\right)^{1/q}. \tag{28}$$

为了证明这个定理,我们先证明

**引理(杨(Young)氏不等式)** 若 $a,b\geqslant 0$, $p,q$ 同上,则

$$ab \leqslant \frac{a^p}{p}+\frac{b^q}{q}. \tag{29}$$

证 考虑 $0<t<+\infty$ 上的函数 $\varphi(t)=t^\alpha-\alpha t$, $0<\alpha<1$. $0<t<1$ 时 $\varphi'(t)>0$, $1<t<\infty$ 时, $\varphi'(t)<0$,所以 $\varphi(t)$ 当 $t=1$ 时,达到最大值,即 $t^\alpha-\alpha t\leqslant 1-\alpha$,而

$$t^\alpha \leqslant \alpha t+(1-\alpha).$$

如果 $a=b=0$,则(29)自然成立,所以我们可以限于 $b\neq 0$ 的情况,令 $\alpha=1/p$,则 $1-\alpha=1-\dfrac{1}{p}=\dfrac{1}{q}$,用 $t=a^p b^{p(1-q)}$,代入上式双方再乘以 $b^q$,即有

$$ab \leqslant \frac{a^p}{p}+\frac{b^q}{q}.$$

**赫尔德不等式的证明** 在上式中令

$a=\left|f(x)\right|\Big/\left(\int_E\left|f(x)\right|^p\mathrm{d}x\right)^{1/p}$, $b=\left|g(x)\right|\Big/\left(\int_E\left|g(x)\right|^q\mathrm{d}x\right)^{1/q}$ (暂设两个分母均不为0),则

$$\int_E a^p\mathrm{d}x=1,\ \int_E b^q\mathrm{d}x=1.$$

以 $a,b$ 代入上式并在 $E$ 上积分即得赫尔德不等式. 若上面至少有一个分母为 0,例如 $\int_E\left|f(x)\right|^p\mathrm{d}x=0$,则 $f(x)=0$ 几乎处处成立而(28)是自明的.

**定理 8(闵可夫斯基(Minkowski)不等式)** 若 $f(x),g(x)\in L^p(E)$, $1<p<+\infty$,则

$$\left(\int_E\left|f(x)+g(x)\right|^p\mathrm{d}x\right)^{1/p}\leqslant\left(\int_E\left|f(x)\right|^p\mathrm{d}x\right)^{1/p}+\left(\int_E\left|g(x)\right|^p\mathrm{d}x\right)^{1/p}. \tag{30}$$

证 在赫尔德不等式中把 $|g(x)|$ 换成 $|f+g|^{p/q}$,即有

$$\int_E\left|f(x)\right|\left|f(x)+g(x)\right|^{p/q}\mathrm{d}x\leqslant\left(\int_E\left|f(x)\right|^p\mathrm{d}x\right)^{1/p}\left(\int_E\left|f(x)+g(x)\right|^p\mathrm{d}x\right)^{1/q}.$$

类似地,把 $|f(x)|$ 换成 $|f+g|^{p/q}$,又有

$$\int_E\left|g(x)\right|\left|f(x)+g(x)\right|^{p/q}\mathrm{d}x\leqslant\left(\int_E\left|g(x)\right|^p\mathrm{d}x\right)^{1/p}\left(\int_E\left|f(x)+g(x)\right|^p\mathrm{d}x\right)^{1/q}.$$

两式相加,注意到 $1+\dfrac{p}{q}=p\left(\dfrac{1}{p}+\dfrac{1}{q}\right)=p$,所以

$$\int_E\left|f+g\right|^p\mathrm{d}x\leqslant\left[\left(\int_E\left|f(x)\right|^p\mathrm{d}x\right)^{1/p}+\left(\int_E\left|g(x)\right|^p\mathrm{d}x\right)^{1/p}\right]\cdot\left(\int_E\left|f(x)+g(x)\right|^p\mathrm{d}x\right)^{1/q}.$$

双方除以 $\int_E\left|f(x)+\right.$

可测函数空间,而如果在定理7,8中这样来理解 $L^{\infty}$,则它们对 $1\leqslant p,q\leqslant +\infty$ 也是成立的.

闵可夫斯基不等式当 $p=2$ 时就是三角形不等式.因此,在 $L^p(E)$ 中我们仍可定义 $f(x)$ 作为一个向量是"长度",称为模或范数,记作

$$\|f\|_{L^p}=\left(\int_E\left|f(x)\right|^p\mathrm{d}x\right)^{1/p},\ 1\leqslant p<+\infty,$$

$$\|f\|_{L^{\infty}}=\operatorname*{esssup}_E|f(x)|.$$

尽管我们不能定义内积因此没有角度、正交性与o.n.系等概念,但有了向量"长度",就可以定义两个向量或两个点 $f,g$ 之距离 $\rho(f,g)=\|f-g\|_{L^p}$.很容易看到,它和 $L^2(E)$ 中的距离一样也具有以下三条基本性质

(i) $\rho(f,g)=\rho(g,f)$;

(ii) $\rho(f,g)\geqslant 0,\rho(f,g)=0\Leftrightarrow f=g$;

(iii) $\rho(f,g)\leqslant\rho(f,h)+\rho(h,g)$.

所以 $L^p(E)$,$1\leqslant p\leqslant +\infty$ 同样也是性质极丰富的空间:它们有距离,也有线性结构,但除非 $p=2$,没有正交性的概念.它们也提供了一系列十分有用的数学框架.不过我们要注意,$L^1$ 和 $L^{\infty}$ 空间有许多重要的性质是一般的 $L^p(1<p<+\infty)$ 空间所没有的.在今后的学习过程中这一点应该牢记.

最后我们要问,是否 $L^p(E)$ 空间也适合里斯-费希尔定理?答案是肯定的:

**定理9**　若 $\{f_k\}$ 是 $L^p$ 空间中按 $\|\cdot\|_{L^p}$ 范数的柯西序列,则必存在 $f\in L^p(E)$ 使在 $L^p$ 意义下 $\lim\limits_{k\to\infty}f_k=f$.

证明略去,我们只提醒一下,$p<\infty$ 时,$L^p$ 意义下的收敛都是以积分表示的一种平均收敛,只有 $L^{\infty}$ 收敛性是除了一个零测度集后的一致收敛性.

# §5　高斯积分

本章前一部分比较系统地介绍了积分理论.我们先讲了从19世纪中黎曼对自古希腊以来的计算面积问题的理论和方法所作的系统的整理,提出了第一个完整的可积性理论.继而介绍了20世纪开始时兴起的勒贝格积分理论.我们指出了,黎曼积分的局限性根源在于它是以只具有有限可加性的若尔当测度为基础的,而勒贝格积分理论则是以具有可数可加性的勒贝格测度为基础的.概率也正是一种具有可数可加性的测度,所以勒贝格积分在概率、统计,以至理论物理中起着不可替代的作用绝非偶然.最后我们则介绍了勒贝格可积函数所形成的一个空间的"阶梯":$\{L^p\}$,$1\leqslant p\leqslant +\infty$,它们提供了重要的解决数学与物理问题的框架.

本章其余部分则介绍一些与积分理论有关的发展.它们都有深刻的、广泛的应用,但又不能说是积分理论的主体的一部分.所以我们的介绍必然更简短一些,可能也显得比较零碎.

这一节里要介绍的高斯积分,从计算的方法来讲实在简单不过,但是它的含意深远.我们先从一个理想化的物理问题开始.

设有一个粒子在 $x$ 轴上运动,粒子的位置限定在 $x=nh$($n$ 为整数)处.如每过一定时段 $\tau$,

粒子就移动一个步长 $h$，到 $x=(n-1)h$，或 $x=(n+1)h$ 处. 如果没有任何物理的理由使这个粒子更易向左或向右走，则我们说粒子从 $nh$ 处移到 $(n-1)h$ 或 $(n+1)h$ 的概率都是 $\frac{1}{2}$. 再过一个时段 $\tau$，粒子又继续随机地左、右移动. 这里一个重要的假设是，粒子每一步是向左还是向右与其过去各步(即其"历史")是无关的. 如果我们记粒子在第 $n$ 步时，即在 $t=n\tau$ 时粒子已从 $x=ih$ 移动到 $x=jh$ 的概率为 $P(ih-jh,n\tau)$，则原来关于粒子运动的概率的假设是

$$P_{ij}=P(ih-jh,\tau)=\begin{cases}\frac{1}{2}, & |i-j|=1,\\ 0, & |i-j|\neq 1.\end{cases} \tag{1}$$

这样一个模型有两个特点：

(i) 齐次性(或称均匀性). 概率只依赖于粒子移动的总的路程，亦即从 $x=ih$ 移动到 $x=jh$ 与从 $x=(i+i_0)h$ 移动到 $x=(j+i_0)h$ 的概率是相同的. 所以 $P_{ij}=P((i-j)h,\tau)$ 是 $i-j$ 的函数.

(ii) 各向同性. 从 $x=ih$ 移动到 $x=jh$ 的概率与从 $x=jh$ 移动到 $x=ih$ 的概率相同. 因此 $P((i-j)h,\tau)$ 其实是 $|i-j|$ 的函数，于是 $P_{ij}=P_{ji}$.

这样建立起来的数学模型称为 1 维随机游动(random walk). 现在我们来计算 $P((i-j)h,n\tau)$. 它是设在初始时刻粒子位于 $x=0$ 处而在经过 $n\tau$ 后移动到 $(i-j)h$ 处的概率. 一共游动了 $n$ 步，如果其中有 $k$ 步是向右，则有 $n-k$ 步是向左，结果则游动了

$$kh-(n-k)h=(-n+2k)h=(i-j)h.$$

如果 $i-j$ 和 $n$ 都固定，则由上式 $k$ 可以唯一地确定. 问题只在于一旦固定了 $k$，则游动结果必达到 $(i-j)h$. 这 $n$ 步中总有 $k$ 步是向右，但是是哪 $k$ 步是不一定的. 这些向右移动的步子的选法共有 $\binom{n}{k}$ 种，而每一步向右、向左的概率又各为 $\frac{1}{2}$，所以 $n$ 步后达到 $x=(i-j)h$ 的概率是 $\frac{1}{2^n}\binom{n}{k}$，

$$P((i-j)h,n\tau)=\frac{1}{2^n}\binom{n}{k}. \tag{2}$$

我们可以从一个树形图(图 4-5-1)来看为什么是这个结果：要想在 $t=n\tau$ 时达到 $P_k$，必须在 $t=(n-1)\tau$ 达到点 $P_{k-1}$ 或点 $P_k$，从点 $P_{k-1}$ 再向前游动有两个可能，到达点 $P_k$ 的概率只有 $\frac{1}{2}$，从点 $P_k$ 向后也是一样，所以

$$P((i-j)h,n\tau)=\frac{1}{2}P((i-j)h-h,(n-1)\tau)+\frac{1}{2}P((i-j)h,(n-1)\tau). \tag{3}$$

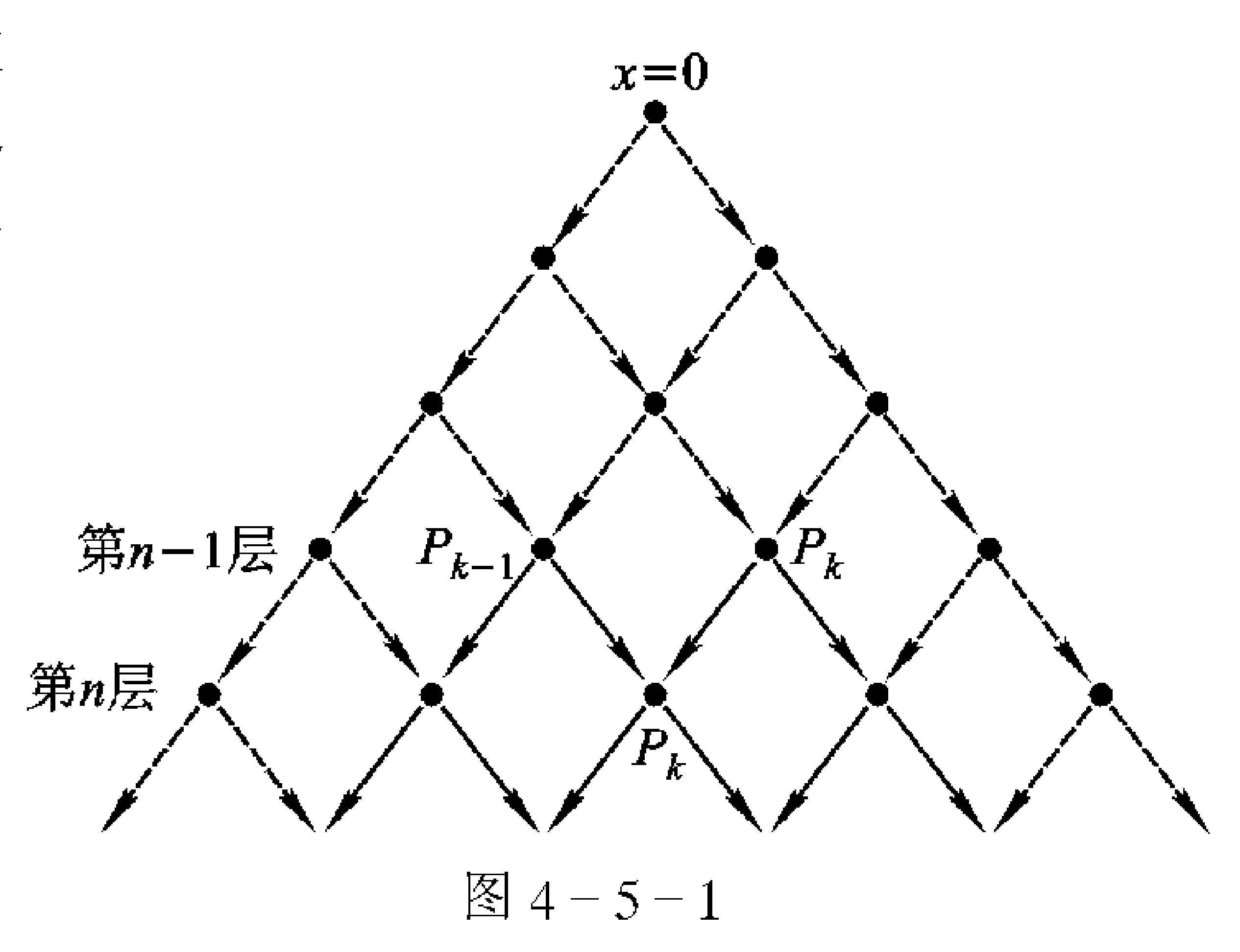


图 4-5-1

但这就是著名的杨辉恒等式

$$\binom{n}{k}=\binom{n-1}{k}+\binom{n-1}{k-1}.$$

若在(3)式中令 $(i-j)h=x$，$(n-1)\tau=t$，就得到

$$P(x,t+\tau)=\frac{1}{2}P(x+h,t)+\frac{1}{2}P(x-h,\tau). \tag{4}$$

这里考虑的是粒子向左与向右的概率同为 $\frac{1}{2}$ 的情况. 如果向右游动与向左游动的概率分别为 $p$ 和 $q=1-p$，而且 $0<p<1$，则代替上式有

$$P(x,t+\tau)=pP(x+h,t)+qP(x-h,\tau).$$

上面画的图的每条水平线 $t=n\tau$ 给出了初始位置为 $x=0$ 的粒子在这一时刻达到各个结点的概率，所以称为一个概率分布. 二项式定理给出了

$$\sum_i P((i-j)h,n\tau)=\sum_k \frac{1}{2^n}\binom{n}{k}=1. \tag{5}$$

(注意，若 $k<0$ 或 $k>n$ 我们规定 $\binom{n}{k}=0$，这样在求和号下就不必具体指出 $k$ 的变化范围.)

(2)式这种概率分布称为二项分布.

上面给出的是一个离散模型. 为了要由宏观量过渡到微观量，就需要令 $h,\tau\to 0$，这样就会得到一个连续模型. 但为此首先要引入概率分布的密度的概念. 如果一个粒子位于长为 $h$ 的某个区间的概率是 $p$，则当区间长 $h\to 0$ 时就会想到该粒子恰好位于某一点上的概率，在合理的设想下它可能为0，这恰好是很容易引起问题的. 因此我们引入概率密度 $q$ 的概念；粒子位于该区间的概率与 $h$ 成正比而为 $qh$，这样，上面的(5)式(不妨令 $j=0$，$n\tau=t$)将成为

$$\sum_i P(ih,t)=1 \Leftrightarrow \int q(x,t)\,\mathrm{d}x=1. \tag{6}$$

现在我们就可以来引入连续模型了. 从(4)式双方减去 $P(x,t)$ 再除以 $\tau$，有

$$\frac{P(x,t+\tau)-P(x,t)}{\tau}=\frac{h^2}{2\tau}\frac{P(x+h,t)-2P(x,t)+P(x-h,t)}{h^2}.$$

再令 $P(x,t)=hq(x,t)$ 即引入概率密度. 令 $\tau,h\to 0$，但 $\frac{h^2}{2\tau}=D$ 不变，即有

$$\frac{\partial q}{\partial t}=D\frac{\partial^2 q}{\partial x^2}. \tag{7}$$

(7)式称为扩散方程，它与第三章§1中引入的热传导方程形式是一样的. 但是它不是从热作为一种流体这种机械—力学的观点出发的，而在本质上是统计的. 它正是爱因斯坦研究布朗运动的基础. 爱因斯坦利用已知的种种物理知识指出了扩散系数 $D$ 与其它物理量的关系，并由此进而得出了阿伏伽德罗(Avogadro)常数——标准状况下一摩尔气体中的分子个数——与从其它方法得到的结果相符得很好，这样才使得原子作为一个实体的存在得到了公认.

现在再回到一个粒子在初始时刻位于 $x=0$ 处的扩散(我们用一维随机游动来描述扩散过程). 初始时刻该粒子位于 $x=0$ 处的概率为1而位于其它位置($x\neq 0$)处的概率为0. 如果用概率密度来表述，若记该概率密度为 $\delta(x)$，则

$$\delta(x)=\begin{cases}\infty, & x=0\\ 0, & x\neq 0,\end{cases}$$

$$\int \delta(x)\mathrm{d}x = 1.$$

正如我们在第二章最后一节所讲的，这个概率密度应该用广义函数来描述，而这个 $\delta(x)$ 正是著名的广义函数——狄拉克(Dirac)测度，或称 $\delta$ 函数. 因此，如果用连续模型来讨论它，我们就需要求解以下的柯西问题

$$\frac{\partial q}{\partial t} = D\frac{\partial^2 q}{\partial x^2}, \qquad q(x,0) = \delta(x). \tag{8}$$

这种以 $\delta(x)$ 为已知数据的解常称为偏微分方程的基本解. 对于扩散方程，基本解是

$$q(x,t) = \frac{1}{2\sqrt{\pi Dt}}\exp\{-x^2/4Dt\}, t>0. \tag{9}$$

这里涉及的广义函数知识在第二章中已讲了一点，下一节再仔细讨论，至于基本解的问题本书已经不可能讲了.

一个概率分布如果其密度是(9)则称为正态分布或高斯分布，因为首先是高斯在讨论观测值的误差分布时遇到了它. 这是一个极其重要的分布，不仅因为它出现在许多似乎全不相似的问题中，还因为概率论中有一个著名的定理(棣莫弗(DeMoivre)-拉普拉斯定理)，大意是说，当二项分布的 $n$ 很大时，可以在一定意义下用正态分布去近似它. 现在我们要提到的是，麦克斯韦在研究气体的分子运动论时也得到了它.

麦克斯韦的论证从现在统计物理学的发展来看有不足之处，然而其基本思想仍然是正确的，而且只需简单的微积分知识就可以了解它.

麦克斯韦考虑在一个容器内处于热平衡状态下的气体分子. 由于其相互碰撞，分子的速度是随机的. 因此我们不能说哪一个确定的分子在给定时刻的准确的速度是多少，而只能问速度在某个给定范围(例如 $x_i$ 方向的速度在 $v_i$ 到 $v_i + \mathrm{d}v_i$ 之间)内的分子数是多少. 如果概率密度是 $f(v)$, $v = (v_1, v_2, v_3) \in \mathbf{R}^3$，则它应该适合以下三个条件：

(i) 规范性条件：

$$\int_{\mathbf{R}^3} f(v)\mathrm{d}v = 1, \mathrm{d}v = \mathrm{d}v_1\mathrm{d}v_2\mathrm{d}v_3,$$

上式表示气体分子的总数总是固定的，不妨设它为 1. 因此 $f(v)$ 必须是 $v$ 的可积函数.

(ii) 统计独立性条件：这就是说，每个分子在各个运动方向(我们暂设有三个独立方向即 $(x_1, x_2, x_3)$，而不讨论分子是否旋转)上速度的分布是互相独立的. 独立性就表示为速度分布(同时考虑三个速度分量)的密度是各个分速度的概率密度函数之乘积.

$$f(v) = f_1(v_1)f_2(v_2)f_3(v_3). \tag{10}$$

(iii) 各向同性条件：即 $f(v)$ 只依赖于 $v^2 = v_1^2 + v_2^2 + v_3^2$：

$$f(v) = \Phi(v_1^2 + v_2^2 + v_3^2). \tag{11}$$

现在我们要证明，只有高斯分布能同时适合以上三个条件. 取(10)与(11)之对数，再对 $v_1$ 求导：

$$\frac{1}{2v_1}\frac{\mathrm{d}}{\mathrm{d}v_1}\ln f_1(v_1) = \frac{\Phi'(v^2)}{\Phi(v^2)}.$$

同理，上式右方也等于 $\frac{1}{2v_2}\frac{\mathrm{d}}{\mathrm{d}v_2}\ln f_2(v_2)$，$\frac{1}{2v_3}\frac{\mathrm{d}}{\mathrm{d}v_3}\ln f_3(v_3)$，因此右方必等于某常数 $A$，而有

$$\Phi(v^2) = B\ \exp(Av^2),$$

$B$ 待定.由可积性条件 $A$ 必须为负:$A = -1/\alpha^2$,$\alpha$ 待定.现在求 $B$,由以上可知应有

$$f_1(\cdot) = f_2(\cdot) = f_3(\cdot) = B^{1/3}\exp(-\cdot/\alpha^2).$$

代入规范性条件有

$$1 = B\left[\int_{-\infty}^{\infty}\exp(-v^2/\alpha^2)\mathrm{d}v\right]^3 = B(\sqrt{\alpha^2\pi})^3,$$

因此 $B^{1/3} = (\alpha^2\pi)^{-1/2}$.而有

$$f_i(v_i) = (\alpha^2\pi)^{-1/2}\exp\{-v_i^2/\alpha^2\}.$$

而概率密度为

$$f(v) = (\alpha^2\pi)^{-3/2}\exp\{-v^2/\alpha^2\}.$$

不过这是一个三维的结果,而上面我们对一维随机游动所给出的结果是一维的.上式中的 $\alpha$ 麦克斯韦指出应为

$$\alpha = \sqrt{2kT/m},$$

$m$ 是分子质量,$T$ 是温度,$k$ 是一个常数,后来称为玻耳兹曼常数.

由麦克斯韦分布

$$f(v) = \left(\frac{m}{2\pi kT}\right)^{3/2}\exp\left(-\frac{1}{2}mv^2/kT\right) \tag{12}$$

可以得出关于气体的一些宏观物理量.就某一个粒子(气体分子)而言,要准确地说出它的某个物理量(例如动能$\frac{1}{2}mv^2$)是不可能的:因为我们只能知道该粒子速度在 $v$ 与 $v+\mathrm{d}v$ 之间的概率.但是要计算大量粒子的某个物理量 $f(v)$ 的平均值是可能的:我们只要把 $f(v)$ 按一定概率分布密度 $\rho(v)$ 求平均即可,$f$ 的平均值或者记为 $\bar{f}$ 或者记为 $\langle f\rangle$:

$$\bar{f} = \langle f\rangle = \int f(v)\rho(v)\mathrm{d}v.$$

(一个量的方差即它与其平均值之偏差之平均值 $\langle[f-\langle f\rangle]^2\rangle$,)平均值也称为该物理量的数学期望值.我们现在就来计算一个气体分子的平均动能.

$$\begin{aligned}\left\langle\frac{1}{2}mv_1^2\right\rangle &= \frac{m}{2}\int v_1^2 f(v_1)\mathrm{d}v \\ &= \frac{m}{2}\left(\frac{m}{2\pi kT}\right)^{1/2}\int_{-\infty}^{\infty} v_1^2\exp\left(-\frac{1}{2}mv_1^2/kT\right)\mathrm{d}v_1.\end{aligned}$$

令 $\left(\frac{m}{2kT}\right)^{1/2}v_1 = t$,有

$$\left\langle\frac{1}{2}mv_1^2\right\rangle = \frac{m}{2}\cdot\frac{1}{\sqrt{\pi}}\cdot\frac{2kT}{m}\int_{-\infty}^{\infty} t^2\exp(-t^2)\mathrm{d}t.$$

很明显,分子在 $x_1$ 方向运动所产生的动能与温度成正比,而且各个方向(即各个自由度)运动所产生的动能均相等,所以这三个自由度所生的动能加起来可以得到

$$\left\langle\frac{1}{2}mv^2\right\rangle = \frac{3}{2}kT.$$

上面这些计算都本质地依赖于计算

$$I_n=\int_0^{+\infty}x^n e^{-x^2}\mathrm{d}x, n=0,1,2,\cdots. \tag{13}$$

计算平均动能时就要用到 $2I_2=\int_{-\infty}^{\infty}x^2e^{-x^2}\mathrm{d}x=\sqrt{\pi}$. 计算方法如下：$n=0$ 时就有著名的高斯积分. $n$ 为其它值时在计算许多函数的数学期望值时有用. 这里的积分限本可取为 $(-\infty,+\infty)$. 但这样一来当 $n$ 为奇数时，因被积函数为奇函数，$(-\infty,+\infty)$ 上之积分为 0. $n$ 为偶数时，被积函数为偶函数，所以 $(-\infty,0)$ 上之积分与 $(0,+\infty)$ 上之积分相等，所以我们只来讨论 $(0,+\infty)$ 上的积分 (13).

先看 $n=0$，利用富比尼定理

$$\iint\limits_{第一象限}e^{-(x^2+y^2)}\mathrm{d}x\mathrm{d}y=\int_0^{+\infty}e^{-x^2}\mathrm{d}x\int_0^{+\infty}e^{-y^2}\mathrm{d}y=I_0^2.$$

左方用极坐标表示为

$$\int_0^{\pi/2}\mathrm{d}\theta\int_0^{+\infty}e^{-r^2}r\mathrm{d}r=\frac{1}{2}\int_0^{\pi/2}\mathrm{d}\theta\int_0^{+\infty}e^{-t}\mathrm{d}t\quad(t=r^2)$$
$$=\frac{\pi}{4}.$$

因此

$$I_0=\frac{\sqrt{\pi}}{2}.$$

计算 $I_n$ 可以用分部积分法. 首先

$$I_1=\int_0^{\infty}xe^{-x^2}\mathrm{d}x=-\frac{1}{2}e^{-x^2}\Big|_0^{\infty}=\frac{1}{2}.$$

然后利用分部积分法. 当 $n\geqslant 2$ 时，注意到 $xe^{-x^2}=-\frac{1}{2}(e^{-x^2})'$，

$$I_n=\int_0^{\infty}x\cdot x^{n-1}e^{-x^2}\mathrm{d}x=-\frac{1}{2}x^{n-1}e^{-x^2}\Big|_0^{\infty}+\frac{n-1}{2}\int_0^{\infty}x^{n-2}e^{-x^2}\mathrm{d}x.$$
$$=\frac{n-1}{2}I_{n-2}.$$

因此

$$I_{2m}=\frac{(2m-1)(2m-3)\cdots 1}{2^{m-1}}I_0=\frac{(2m-1)(2m-3)\cdots 1}{2^m}\sqrt{\pi},$$
$$I_{2m-1}=\frac{1}{2}(m-1)!. \tag{14}$$

上面我们讲到当 $n\to\infty$ 时，二项分布将在某种意义下以高斯分布为极限，但是没有详细说明. 下面我们来看一下物理学家是怎样用近似方法来得到这个结论的，这对于我们了解微积分的基本思想以及如何应用它们是很有益处的. 其中我们还讲到著名的斯特林 (Stirling) 公式，它也是很有用的. 以下材料引自 F. 瑞夫，统计物理学（《伯克利物理学教程》第五卷，附录 A.1，高斯分布，437～443 页，该书由科学出版社 1979 年出版）.

我们讲到一维随机游动时是假定了粒子向左、向右移动的概率均为 $\frac{1}{2}$：而有 (2) 式

$P((i-j)h,n)=\frac{1}{2^n}\binom{n}{k}$. 但若向右与向左游动的概率各为 $p$ 与 $q=1-p$，则应有如下的概率

$$P(k,n)=\binom{n}{k}p^kq^{n-k}=\frac{n!}{k!\ (n-k)!}p^kq^{n-k}. \tag{15}$$

若 $n$ 很大，计算 $n!$ 将是相当困难的. 但是因为 $0<p,q<1$，故当 $k\gg1$ 或 $n-k\gg1$ 时，上述概率仍然很小，而 $P(k,n)$ 极大处 $k=\tilde{k}$ 总在靠近中间的地方(当 $p=q=\frac{1}{2}$ 时恰好在正中. 正如二项定理之系数在“中点”($n$ 为偶数时)或紧靠“中点”左右两点($n$ 为奇数时)最大. 所以当我们想找 $P(k,n)$ 的近似公式时，除非 $p\ll1,q\ll1$，$\tilde{k}$ 总是随着 $n$ 同样取相当大的值. 现在我们就来求 $P(k,n)$ 关于 $k$ 的极大值.

$k$ 是一个整数，所以每一次 $k$ 的增量至少是 1，这本来已是相当大的数了，但是如果 $k$ 本身就已很大，其“增量”1 与 $k$ 相比仍然可以认为是“无穷小”，而且它所引起的 $P$ 的变化也是极小的(这是很值得注意的事. 在物理学家看来，甚至不太小的数也可以看作是无穷小. 所以，一旦离开思辨的天地而进入现实，确有“海阔天空”之感. 但是，这是不是说以前讲的数学都是没有意义了呢？暂时不要去争论，而是要多向物理学家学一点东西)：

$$|P(k+1,n)-P(k,n)|\ll P(k,n). \tag{16}$$

所以我们不妨认为 $k$ 是连续变量而 $P(k,n)$ 是它的光滑函数. 我们又知道 $\ln P$ 比之 $P$ 是变化更缓慢的函数，因此，取(15)之对数

$$\ln P=\ln n!\ -\ln k!\ -\ln(n-k)!\ +k\ln p+(n-k)\ln q, \tag{17}$$

而且双方对 $k$ 取导数：这里我们要注意，因为 $k$ 是一个很大的整数，其增量 $\Delta k=1$ 可以认为是“无穷小”，从而可以认为 $k$ 是连续变量，并且可以用 $\frac{\Delta\ln k!}{\Delta k}$ 代替 $\frac{\mathrm{d}\ln k!}{\mathrm{d}k}$，于是

$$\begin{aligned}\frac{\mathrm{d}\ln k!}{\mathrm{d}k}&\approx\frac{\ln(k+\Delta k)!\ -\ln k!}{\Delta k}=\frac{\ln(k+1)!\ -\ln k!}{1}\\&=\ln\frac{(k+1)!}{k!}=\ln(k+1)\approx\ln k.\end{aligned}$$

所以，由(17)对 $k$ 求导即有

$$\begin{aligned}\frac{\mathrm{d}\ln P}{\mathrm{d}k}&\approx-\ln k+\ln(n-k)+\ln p-\ln q\\&=\ln\frac{n-k}{k}\cdot\frac{p}{q}.\end{aligned} \tag{18}$$

为了求 $\ln P$ 之极大值应该令上式为 0，即有

$$k=\tilde{k}=np\qquad(注意\ p+q=1).$$

现将 $\ln P$ 在 $k=\tilde{k}$ 处展为泰勒级数，注意到 $\left.\frac{\mathrm{d}\ln P}{\mathrm{d}k}\right|_{k=\tilde{k}}=0$，即得

$$\ln P=\ln P(\tilde{k},n)+\frac{1}{2!}\frac{\mathrm{d}^2\ln P(\tilde{k},n)}{\mathrm{d}\tilde{k}^2}(k-\tilde{k})^2,$$

这里我们略去了 $(k-\tilde{k})^3$ 以及更高阶的量. 为了计算 $\frac{\mathrm{d}^2\ln P}{\mathrm{d}k^2}$，我们对(18)再对 $k$ 求导一次，有

$$\frac{\mathrm{d}^2\ln P(\tilde{k},n)}{\mathrm{d}\tilde{k}^2} = -\frac{1}{\tilde{k}} - \frac{1}{n-\tilde{k}} = \frac{n}{-\tilde{k}(n-\tilde{k})},$$
$$= -\frac{1}{npq}.$$

所以可得 $\ln P$ 的近似式

$$\ln P(k,n) \approx \ln P(\tilde{k},n) - \frac{(k-\tilde{k})^2}{2npq},$$
$$P(k,n) = \tilde{P}\exp[-(k-\tilde{k})^2/2npq], \tag{19}$$

这里 $\tilde{P} = P(\tilde{k},n)$. 为了求它的值,我们要利用规范化条件,即全部概率(从 $k=0$ 到 $k=n$)之和为 1:

$$\sum_{k=0}^{n} P(k,n) = \sum_{k=0}^{n} P(k,n)\cdot\Delta k, \quad \Delta k = 1.$$

和上面一样的想法使我们把上式看成一个积分和,而有

$$\int P(k,n)\mathrm{d}k = 1.$$

注意到 $\tilde{k} = np$,则得 $P(k,n)$ 之近似式为

$$P(k,n) = \frac{1}{\sqrt{2\pi npq}}\mathrm{e}^{-(k-np)^2/2npq}. \tag{20}$$

这正是高斯分布,所以我们说高斯分布是二项分布的极限情况.

上面我们讲到如何近似计算一个大整数 $n$ 之阶乘 $n!$. 实际上利用类似的想法可以得到一个很方便的近似式. 我们有

$$\ln n! = \ln 1 + \ln 2 + \cdots + \ln n = \sum_{k=1}^{n} \ln k.$$

和上面一样,我们认为每一项都有一个因子 $1 = \Delta k$,而把上式看成一个积分和,这样就可得出一个近似式.

$$\ln n! \approx \int_1^n \ln x\mathrm{d}x = [x \ln x - x]_1^n = n \ln n - n + 1.$$

所以当 $n>1$ 时,略去上式最后一项即有

$$\ln n! \approx n\ln n - n. \tag{21}$$

但是这还不是最好的近似. 更好的还有著名的斯特林公式:

$$\ln n! \approx n\ln n - n + \frac{1}{2}\ln(2\pi n), \tag{22}$$

或

$$n! \approx \mathrm{e}^{-n}\sqrt{2\pi}n^{n+\frac{1}{2}}. \tag{23}$$

这里的≈表示两边的比值当 $n\to\infty$ 时趋于 1. 当 $n\gg 1$ 时,因为 $\ln n$ 比之 $n\ln n$ 是低阶量,可以略去,这样就得到(21)式.

当然读者会感到这里的讲法并不严格,但其主要思想是正确的. 读者如能找到较详尽的数学分析教材中关于斯特林公式的证明以及概率论教材中关于棣莫弗–拉普拉斯定理的证明就可以

看到,主要思想是相同的.但是关于无穷大、无穷小量的估计在那些书中做得细致得多.

最后我们讲一下高斯积分的一个简单的推广.高斯积分主要特点是被积函数是一个指数函数,而其指数是一个负的二次式.因此,我们不妨考虑一个 $n$ 重积分

$$I=\int_{\mathbf{R}^n} e^{-\langle Ax,x\rangle}\,dx, \tag{24}$$

这里 $A$ 是一个正定的 $n$ 阶对称矩阵:$A=(a_{ij})$.规定 $A$ 为正定的原因在于当 $|x|\to\infty$ 时,$-\langle Ax,x\rangle\to-\infty$ 而 $e^{-\langle Ax,x\rangle}$ 是可积的.对称矩阵一定可以用正交变换化为对角形,即存在一个正交矩阵 $T({}^tT\cdot T=id)$ 使

$$T^{-1}AT=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix},$$

这里 $\lambda_1,\cdots,\lambda_n$ 是 $A$ 的特征根.因为 $A$ 是对称矩阵,故其特征根均为实数.而当我们设 $A$ 为正定时,$\lambda_1,\lambda_2,\cdots,\lambda_n>0$.现在作变量变换

$$x=Ty,$$

则

$$\begin{aligned}\langle Ax,x\rangle &= \langle ATy,Ty\rangle=\langle {}^tTATy,y\rangle\\ &=\langle T^{-1}ATy,y\rangle=\sum_{i=1}^n\lambda_i y_i^2.\end{aligned}$$

又因正交矩阵之行列式等于 $\pm1$,故

$$dx=|\det T|\,dy=dy.$$

代入(24)即有

$$\begin{aligned}I &= \int_{\mathbf{R}^n}\exp\Big(-\sum_{i=1}^n\lambda_i y_i^2\Big)dy=\prod_{i=1}^n\int_{-\infty}^{\infty}e^{-\lambda_i y_i^2}dy_i\\ &=\prod_{i=1}^n(\pi/\lambda_i)^{\frac{1}{2}}.\end{aligned} \tag{25}$$

如果 $A$ 不是实矩阵,而只有 $\mathrm{Re}A>0$,也有相应的结果.因为需要更多的预备知识,我们只好从略.

# §6　分部积分法、广义函数、索伯列夫(Sobolev)空间

**1.分部积分法与对偶性**　黎曼积分中起十分重要作用的分部积分法在勒贝格积分理论中是否适用?答案是肯定的,但是对函数的可微性、可积性要有更细致的规定.分部积分法之重要在于,它体现了一个非常重要的数学概念——对偶性——所以我们宁可暂时不问是那一类积分,也不把导数看成是 $\dfrac{\Delta f}{\Delta x}$ 的极限而对它加以推广,以便突出分部积分法的主要思想.为此先按我们"常见的"形式看一下分部积分公式,但是是在 $n$ 维空间的一个开集 $\Omega$ 上去考察它.先设 $\partial\Omega$ 是 $x_i=0$,则

$$\int_{\Omega}\frac{\partial f}{\partial x_i}(x)\varphi(x)\mathrm{d}x=\int_{\partial\Omega}\varphi f\Big|_{x_i=0}\mathrm{d}x'-\int_{\Omega}f(x)\frac{\partial\varphi}{\partial x_i}(x)\mathrm{d}x,$$

$x'=(x_1,\cdots,\hat{x}_i,\cdots,x_n)$. $x_i$ 上面加了一个 ^,表示从 $x$ 的坐标中把 $x_i$ 删去.我们希望 $\partial\Omega$ 上的积分不出现,而这就需要例如假设

$$\varphi|_{\partial\Omega}=0.$$

但是仔细想一下,问题并不那么简单,因为上式是 $n$ 重积分,我们又用了富比尼积分把它化为逐次积分……等等.为此,对 $\Omega$ 的边缘 $\partial\Omega$ 的光滑性就要有所限制.一维情况则简单多了,因为 $\mathbf{R}^1$ 中的开集最多是可数多个区间的并,所以不妨设 $\Omega$ 就是区间 $(a,b)$,而 $\partial\Omega$ 就是两个点 $a$ 和 $b$.为了回避这个问题,我们设 $\varphi(x)\in C_0^k(\Omega)$,下标 0 表示 $\varphi(x)$ 的支集是 $\Omega$ 的紧子集.紧集就是有界闭集,而因 $\varphi(x)$ 只定义于 $\Omega$ 上而不是 $\bar{\Omega}$ 上,所以 supp $\varphi(x)$ 是 $\Omega$ 的一个紧子集,见图 4-6-1.由它到 $\partial\Omega$ 一定有一个正的距离,所以 $\varphi(x)$ 在 $\partial\Omega$ 附近为 0(图 4-6-1).

所以若补充定义 $\varphi(x)$ 在 $\Omega$ 外为 0,则其光滑性不变,即有 $\varphi(x)\in C_0^k(\mathbf{R}^n)$,而 $\partial\Omega$ 可以放在 supp$\varphi$ 外任何位置,可以任意变形而不影响积分.于是我们有

$$\int_{\Omega}\frac{\partial f}{\partial x_i}(x)\varphi(x)\mathrm{d}x=-\int_{\Omega}f(x)\frac{\partial\varphi(x)}{\partial x_i}\mathrm{d}x. \tag{1}$$

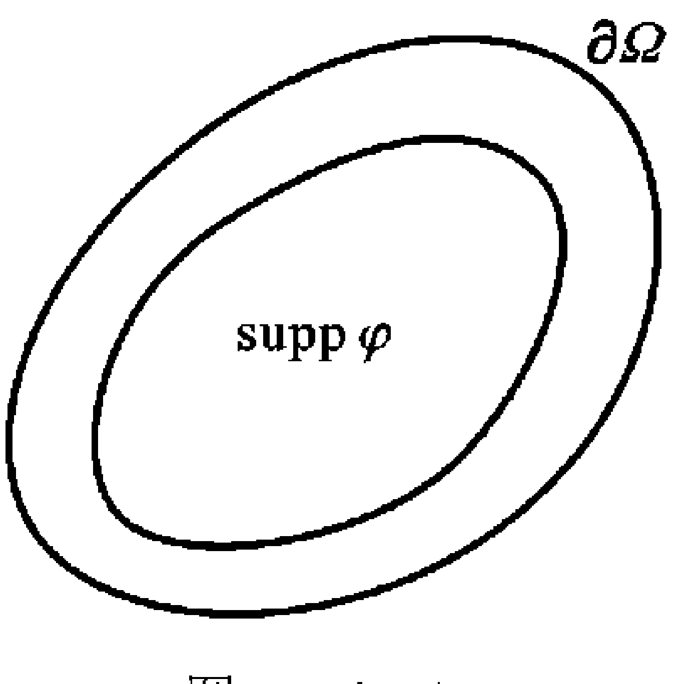


图 4-6-1

如果固定 $f(x)$,而令 $\varphi(x)\in C_0^k$ 变动,则上式双方都是 $\varphi(x)$ 的线性泛函,$f(x)$ 与 $\varphi(x)$ 处于互相对偶的地位:左方是对 $f$ 求导,转到右方则变成了对 $\varphi$ 求导.$f$ 是生成泛函之元,$\varphi$ 是泛函作用于其上之元;(1)式指出,对前者施以 $\frac{\partial}{\partial x_i}$ 就对应于对后者施以 $-\frac{\partial}{\partial x_i}$;$f$ 与 $\varphi$ 是对偶的,$\frac{\partial}{\partial x_i}$ 与 $-\frac{\partial}{\partial x_i}$ 也是对偶的,分部积分法就建立了这样一个对偶关系.这就是我们所说的"对偶性"的一个侧面.

现在的问题是,即令 $f(x)$ 不可求导,但若能够找到另一个 $g_i(x)$ 适合

$$\int_{\Omega}g_i\varphi(x)\mathrm{d}x=-\int_{\Omega}f(x)\frac{\partial\varphi(x)}{\partial x_i}\mathrm{d}x,$$

则 $g_i(x)$ 起了 $\frac{\partial f}{\partial x_i}$ 的作用,因此可以把它看作 $f$ 的某种广义的导数.尽管它与 $\frac{\Delta f}{\Delta x_i}$ 有何关系并不清楚.$f(x)$,$g_i(x)$ 要求某种可积性,不过也很清楚不需要它们在 $\Omega$ 上可积,因为实际上积分是在 $\Omega$ 的紧子集 supp$\varphi$ 上进行的.而因当 $\varphi$ 变动时 supp$\varphi$ 也在变,所以指定 $f(x)$,$g_i(x)$ 在 $\Omega$ 的某个固定紧子集上可积也不行,所以我们给出

**定义 1**　若 $f(x)$ 在 $\Omega$ 之任一紧子集上属于 $L^p$,就说 $f(x)$ 局部 $p$ 幂可积(局部 $L^p$),记作 $f\in L_{\mathrm{loc}}^p(\Omega)$.$1\leqslant p\leqslant+\infty$.

**定义**

我们可以仿此定义高阶($m$ 阶)广义导数,为此,需设 $\varphi(x)\in C_0^m(\Omega)$,而有

**定义 2′** 设 $f(x)$ 与 $g_\alpha(x)\in L^1_{\mathrm{loc}}(\Omega)$,若对一切 $\varphi(x)\in C_0^m(\Omega)$ 均有

$$\int_\Omega g_\alpha(x)\varphi(x)\mathrm{d}x=(-1)^{|\alpha|}\int_\Omega f(x)\frac{\partial^\alpha\varphi(x)}{\partial x^\alpha}\mathrm{d}x,\tag{3}$$

则称 $g_\alpha(x)$ 为 $f(x)$ 之 $m$ 阶广义导数(弱导数),记作 $g_\alpha(x)=\dfrac{\partial^\alpha f}{\partial x^\alpha}$. 以上 $\alpha=(\alpha_1,\cdots,\alpha_n)$ 为重指标,且 $|\alpha|=\alpha_1+\cdots+\alpha_n=m$.

若对适合 $|\alpha|\leqslant m$ 的一切 $\alpha$,$f(x)$ 都具有弱广义导数 $\dfrac{\partial^\alpha f}{\partial x^\alpha}$,则称 $f(x)$ 为 $m$ 阶广义可微(弱可微).

由于 $L^1_{\mathrm{loc}}(\Omega)$ 函数都是几乎处处定义的,(弱)广义导数也都只是几乎处处定义的.

下面举一些例子,这里均设 $\Omega=(-1,1)\subset\mathbf{R}^1$.

**例 1** 令 $f(x)=|x|$,很明显,按古典定义 $f(x)$ 除在 $x=0$ 处以外均有导数:这里我们利用了 $\varphi\in C_0^1(\Omega)$ 所以 $\varphi(1)=\varphi(-1)=0$.

$$\frac{\mathrm{d}f}{\mathrm{d}x}=\begin{cases}1, & 0<x<1,\\ -1, & -1<x<0\end{cases}$$

也是 $f$ 的广义导数. 事实上,若取 $\varphi(x)\in C_0^1(\Omega)$ 则

$$\begin{aligned}\int_{-1}^1\frac{\mathrm{d}f}{\mathrm{d}x}\varphi(x)\mathrm{d}x&=-\int_{-1}^0\varphi(x)\mathrm{d}x+\int_0^1\varphi(x)\mathrm{d}x\\&=-x\varphi(x)\Big|_{-1}^0+\int_{-1}^0 x\varphi'(x)\mathrm{d}x\\&\quad+x\varphi(x)\Big|_0^1-\int_0^1 x\varphi'(x)\mathrm{d}x=-\int_{-1}^1|x|\varphi'(x)\mathrm{d}x.\end{aligned}$$

所以 $\dfrac{\mathrm{d}f}{\mathrm{d}x}$ 也是 $|x|$ 的(弱)广义导数.

**例 2** 令 $H(x)$ 为赫维赛德函数:

$$H(x)=\begin{cases}1 & x\geqslant 0,\\ 0 & x<0.\end{cases}$$

它不会有(弱)广义导数. 实际上,若 $g(x)\in L^1_{\mathrm{loc}}(R)$ 是 $H(x)$ 的广义导数,对 $x_0>0$ 取 $\varphi(x)$ 使其支集在 $(x_0-h,x_0+h)\subset(0,1)$ 中,应有

$$\int_{x_0-h}^{x_0+h}H(x)\varphi'(x)\mathrm{d}x=-\int_{x_0-h}^{x_0+h}g(x)\varphi(x)\mathrm{d}x,$$

但式左显然为 0,所以式右对适合以上条件的 $\varphi(x)$ 必为 0. 用后面讲的磨光方法即可证明 $g(x)=0$ 在 $x>0$ 处几乎处处成立. 同样可知 $g(x)=0$ 在 $x<0$ 处几乎处于处处成立. 总之,若 $H(x)$ 有一个局部可积函数 $g(x)$ 为其(弱)广义导数,则 $g(x)=0$ 几乎处处成立. 于是由广义导数的定义,对任意 $\varphi(x)\in C^1(-1,1)$,

$$0=\int_{-1}^1 g(x)\varphi(x)\mathrm{d}x=-\int_{-1}^1 H(x)\varphi'(x)\mathrm{d}x=-\int_0^1\varphi'(x)\mathrm{d}x=\varphi(0).$$

所以就会得出结论:任意 $\varphi(x)\in C_0^1((-1,1))$ 都在 $x=0$ 处为 0. 这个错误结论说明赫维赛德函数没有(弱)广义导数. 它在广义函数意义下却有导数:广义函数 $\delta(x)$.

**例 3** 把例 1、例 2 联合起来看，我们发现

$$\frac{\mathrm{d}}{\mathrm{d}x}(|x|)=2H(x)-1.$$

所以$\frac{\mathrm{d}^2}{\mathrm{d}x^2}(|x|)$不能是定义 2 中讲的广义导数，这就告诉我们，沿着对偶性的思想走下去，必然会越出至今我们讨论过的领域. 第二章 §4 已经指出，古典的函数定义在刻画物理世界上是不够用的，例 2 中的 $H(x)$之"导数"应该是 $\delta$ 函数，而它在物理上是十分自然而又有用的. 它甚至超出了广义导数的范围，对此我们将在下面详述.

从这几个例子看到，(弱)广义导数确实超越了古典的导数概念，那么自然会问，$f(x)$的古典意义的导数是否一定是它的(弱)广义导数呢？不一定. 因为(弱)广义导数是由分部积分法衍生出来的，如果 $f(x)$的古典意义下的导数能满足分部积分法的要求，则它自然就是(弱)广义导数，否则就不一定了. 分部积分法的要求其实就是关于一个函数的可积性与其原函数的关系问题. 在 §1 中我们已看到，在黎曼积分理论中，这不是一个简单问题，在勒贝格积分理论中二者的关系如何实际上也非常复杂. 所以我们现在走的是另外一条路. 我们不去考虑例如(弱)广义导数是否与古典意义导数几乎处处相等；$\frac{\Delta f}{\Delta x}$是否几乎处处收敛于(弱)广义导数；……我们现在考虑的是如何沿着对偶性的路走下去.

为此，我们首先注意到 $\varphi(x)$的选取. 我们要求 $\varphi(x)\in C_0^k(\Omega)$，因此，$\varphi(x)$在$\partial\Omega$ 的附近恒为 0，而不仅只在$\partial\Omega$ 上为 0. 这样做其实有两个原因，其一我们已经讲过了，即避免了分部积分法中必然会遇到的$\partial\Omega$ 之光滑性问题. 另一方面是以后我们常要考虑函数序列收敛的问题. 如果 $C_0^k(\Omega)$中的序列$\{\varphi_l(x)\}$在 $\bar{\Omega}$上一致收敛，其极限函数当然在$\partial\Omega$ 上为 0，但不一定在$\partial\Omega$ 某个邻域中为 0；如果$\{\varphi_l(x)\}$在 $L^p$ 中收敛，则极限函数是否在$\partial\Omega$ 上为 0 也不得而知. 再者我们要注意 $k$ 的选取. 如果我们只需讨论一阶导数，则 $k=1$ 即足够应用，但如果求导阶数不确定，则最明智的选取是令 $k=\infty$，而考虑 $\varphi(x)\in C_0^\infty(\Omega)$. 但这里又产生了一个新问题. 例如对一阶导数 $\frac{\partial f}{\partial x_i}$，我们希望某个式子对 $\varphi\in C_0^1(\Omega)$成立，此式当然对 $\varphi\in C_0^\infty(\Omega)\subset C_0^1(\Omega)$也成立. 但若我们只知道上式对 $\varphi\in C_0^\infty(\Omega)$成立，怎样保证它对 $\varphi(x)\in C_0^1(\Omega)$也成立呢？办法在于我们可以用一串$\varphi_m(x)\in C_0^\infty(\Omega)$逼近 $\varphi(x)\in C_0^1(\Omega)$，使得$\frac{\partial\varphi_m}{\partial x_i}$在 $\bar{\Omega}$上一致收敛于$\frac{\partial\varphi}{\partial x_i}$. 上式当 $\varphi=\varphi_m$ 时是成立的. 令 $m\to\infty$即知它对任意 $C_0^1(\Omega)$函数 $\varphi(x)$也成立. $\{\varphi_m\}$的作法下面讲磨光算子时会看到. 所以，下面我们讲的 $\varphi(x)$都选自 $C_0^\infty(\Omega)$，对它们还要规定一个很特殊的收敛性. $C_0^\infty(\Omega)$在赋以这种收敛性后称为试验函数空间. $\varphi(x)$称为试验函数. 不过要注意，我们会需要许多种不同的试验函数空间. 上述的 $C_0^\infty(\Omega)$是最常见的一种，我们以后恒记为 $D(\Omega)$.

现在要详细讨论 $D(\Omega)$. $D(\Omega)$的元素有很好的性质：它们可以任意求导(但是正如第三章中讲的，它们决非解析函数). 它们有很多用处，第三章 §4 中讲到如何利用它们来作"单位分解"——这是一个用于化整体问题为局部问题的有力工具，还讲到磨光算子. 现在我们就来详细讨论磨光算子，仍然采用第三章 §4 的记号. 令 $x,y\in\mathbf{R}^n$，而

$$j_\delta(x-y)=\begin{cases}\exp\left\{\left(\dfrac{|x-y|^2}{\delta^2}-1\right)^{-1}\right\}, & |x-y|\leqslant\delta,\\ 0, & |x-y|>\delta.\end{cases} \tag{4}$$

很明显 $j_1(x)$ 是一个支集在球 $|x|\leqslant 1$ 中的 $C_0^\infty(\mathbf{R}^n)$ 函数，而 $j_\delta(x-y)=j_1\left(\dfrac{x-y}{\delta}\right)$. 令

$$L=\int_{\mathbf{R}^n} j_1(x)\mathrm{d}x,$$

则 $0<L<+\infty$，而且 $\int_{\mathbf{R}^n} j_\delta(x-y)\mathrm{d}x=L\delta^n$. 我们称(4)为一个磨光核，而称以下的算子为磨光算子：

$$M_\delta(f)(x)=\frac{1}{L\delta^n}\int_{\mathbf{R}^n} f(y)j_\delta(x-y)\mathrm{d}y. \tag{5}$$

可以很容易地证明

**定理 1** 设 $f\in L^1_{\mathrm{loc}}(\mathbf{R}^n)$，则

(i) $M_\delta(f)(x)\in C^\infty(\mathbf{R}^n)$.

(ii) 若进一步设 $\operatorname{supp} f=K\subset\subset\mathbf{R}^n$ ("$\subset\subset$"表示左方为右方之紧子集，这个记号也称"紧包含")则 $M_\delta(f)$ 也有紧支集于 $K$ 之 $\delta$ 邻域 $K_\delta=\bigcup\limits_{x\in K}B_\delta(x)$ ($B_\delta(x)$ 表示以 $\delta$ 为半径的 $n$ 维球体)内.

(iii) 若 $f$ 在 $\Omega$ 中连续，则 $M_\delta(f)$ 在 $\Omega$ 之任一紧子集中一致收敛于 $f$ ($\delta\to 0$).

(iv) 若 $f\in L^p(\mathbf{R}^n)$，$1\leqslant p<+\infty$，则在 $L^p(\mathbf{R}^n)$ 内有 $M_\delta(f)\to f$ ($\delta\to 0$).

证 (i)就是经典的在积分号下求导的定理.

(ii) 令 $x$ 在 $K_\delta$ 之外，则对任意 $y\in K$，$x\notin B_\delta(y)$ 而有 $|x-y|\geqslant\delta$，(5)式实际上是在 $K$ 上的积分，而有 $j_\delta(x-y)=0$ ($x$ 在 $K_\delta$ 之外)，因而 $M_\delta(f)(x)=0$.

(iii) 因为 $\dfrac{1}{L\delta^n}\int_{\mathbf{R}^n} j_\delta(x-y)\mathrm{d}y=1$，可见

$$|M_\delta(f)(x)-f(x)|\leqslant\frac{1}{L\delta^n}\int_{\mathbf{R}^n}|f(y)-f(x)|j_\delta(x-y)\mathrm{d}y.$$

这里我们用到了 $j_\delta(x-y)\geqslant 0$. 令 $x-y=z$，上式右方为

$$\frac{1}{L\delta^n}\int_{\mathbf{R}^n}|f(x-z)-f(x)|j_\delta(z)\mathrm{d}z,$$

这里的积分区域实际上是 $|z|\leqslant\delta$. 当 $x$ 在 $\Omega$ 之任一个紧子集 $K$ 内时，若 $\delta$ 充分小，$x-z\in K_\delta\subset\Omega$. 因为 $f$ 在 $\Omega$ 中连续，故必在 $K_\delta$ 上一致连续，从而当 $|z|\leqslant\delta$ 充分小时 $|f(x-z)-f(x)|<\varepsilon$，$\varepsilon$ 是任意小正数，由此命题得证.

(iv) 因为 $f\in L^p(\mathbf{R}^n)$，而每一个 $L^p$

$$|f(x-y)|j_\delta(y)=|f(x-y)|(j_\delta(y))^{1/p}(j_\delta(y))^{1/q},$$

对内层积分应用赫尔德不等式知

$$\| M_\delta(f)(x)\|_{L^p}\leqslant\frac{1}{L\delta^n}\left\{\int_{\mathbf{R}^n}\mathrm{d}x\left[\int_{\mathbf{R}^n}j_\delta(y)\mathrm{d}y\right]^{p/q}\left[\int_{\mathbf{R}^n}\left|f(x-y)\right|^p j_\delta(y)\mathrm{d}y\right]\right\}^{1/p}.$$

因为 $\int_{\mathbf{R}^n}j_\delta(y)\mathrm{d}y=L\delta^n$，代入上式有

$$\begin{aligned}\| M_\delta(f)(x)\|_{L^p}&\leqslant\left(\frac{1}{L\delta^n}\right)^{1/p}\left[\int_{\mathbf{R}^n}\mathrm{d}x\int_{\mathbf{R}^n}\left|f(x-y)\right|^p j_\delta(y)\mathrm{d}y\right]^{1/p}\\&=\left(\frac{1}{L\delta^n}\right)^{1/p}\left[\int_{\mathbf{R}^n}j_\delta(y)\mathrm{d}y\int_{\mathbf{R}^n}\left|f(x-y)\right|^p\mathrm{d}x\right]^{1/p}\\&=\| f\|_{L^p}.\end{aligned}$$

这里应用了富比尼定理. 这就是说，磨光算子将 $L^p(\mathbf{R}^n)$ 映入 $L^p(\mathbf{R}^n)$ 而且范数不增. 因

$f(x)\in L^p(\mathbf{R}^n)$，先作 $g(x)\in C_0(\mathbf{R}^n)$，使 $\| f-g\|_{L^p}<\frac{\varepsilon}{3}$，于是我们有

$$\begin{aligned}\| M_\delta(f)-f\|_{L^p}\leqslant&\| M_\delta(f)-M_\delta(g)\|_{L^p}+\| M_\delta(g)-g\|_{L^p}\\&+\| g-f\|_{L^p}.\end{aligned}$$

第一项与第三项均小于 $\varepsilon/3$，对于第二项，因为 $g(x)\in C_0(\mathbf{R}^n)$，利用(iii)即知，当 $\delta$ 充分小
小于$\varepsilon/3$，定理证毕.

我们附带提到，若 $f\in L^p_{\mathrm{loc}}(\mathbf{R}^n)$，以上讨论仍成立.

不过，上面的讨论当 $p=+\infty$时失效.

现在回到例 2，在那里我们说，若 $g(x)\in L^1_{\mathrm{loc}}(-1,1)$ 且对任意 $\varphi(x)\in C^1_0(-1,$
$\int_{x_0-h}^{x_0+h}g(x)\varphi(x)\mathrm{d}x=0$，则必在$(x_0-h,x_0+h)$中，有 $g(x)=0$ 几乎处处成立. 现在补一

$\frac{\partial f}{\partial x_i}$时，是把(2)式左方作为一个作用在 $C_0^\infty(\Omega)$上的线性泛函来定义的.凡用线性泛函来处理一个数学问题时，我们常用“弱”这个限制词.与此相对，若用范数来讨论问题时则常用“强”字.因此我们自然会问，有没有强广义导数的概念？有，而且这个概念有很深刻的物理背景.第三章变分法一节的最后，我们讲到从黎曼的时代，人们就知道用能量积分

$$I(u)=\iint_\Omega\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u^2}{\partial y}\right)\right]\mathrm{d}x\mathrm{d}y$$

来刻画静电场，而且指出“真正的”静电场必使上述能量积分达到最小值.由此就提出了“极小化”序列$\{u_n\}$，即 $u_n\in C^1(\bar{\Omega})$，而 $I(u_n)\to\min$ 这个思想.但是如果限制在黎曼积分框架下，$\{u_n\}$以及$\left\{\frac{\partial u_n}{\partial x}\right\}$都不一定有极限存在，正是因此，我们才看到勒贝格积分理论何以优于黎曼积分理论：至少我们可以设想$\{u_n\}$与$\left\{\frac{\partial u_n}{\partial x}\right\}$在$L^2$ 意义下有极限 $\bar{u}$与 $\bar{v}$，因此有理由把 $\bar{v}$看作 $\bar{u}$在某种意义下的导数，这样我们就引出了

**定义 3**　设 $f(x)\in L^p(\Omega)$，若存在一个 $C^1(\Omega)$函数序列$\{f_k(x)\}$使得$\|f_k-f\|_{L^p}\to 0$，而且$\frac{\partial f_k}{\partial x_i}\in L^p(\Omega)$有 $L^p$ 意义下的极限$g_i(x)\in L^p(\Omega)$：$\left\|\frac{\partial f_k}{\partial x_i}-g_i\right\|_{L^p}\to 0$，则称 $g_i$ 为$f$ 对于 $x_i$ 的(强)广义导数并仍记以$\frac{\partial f}{\partial x_i}=g_i$.

仿此，可以定义高阶的(强)广义导数.

于是产生一个问题：强弱广义导数是否一致？答案是肯定的，这是一个重要的结论.

**定理 2**　设 $f(x)\in L_{\mathrm{loc}}^p(\Omega)$而且具有强广义导数$\frac{\partial f}{\partial x_i}=g_i(x)\in L_{\mathrm{loc}}^p(\Omega)$，则 $g_i(x)$也是$f(x)$之弱广义导数.若将强弱二字对调，结论仍成立.以上 $1\leqslant p<+\infty$.

**证**　前面讲弱广义导数定义时恒取 $p=1$，实际上，取 $1\leqslant p<+\infty$也可以，这一点下面不再赘言.

(强)⇒(弱).任取 $\varphi(x)\in C_0^1(\Omega)$，则由强广义导数定义，先对 $f_k(x)$使用分部积分法(这是许可的，因为 $f_k(x)\in C^1(\Omega)$)然后再在积分号下求极限(这里可以应用赫尔德不等式)，于是有

$$\int_\Omega f_k(x)\frac{\partial\varphi(x)}{\partial x_i}\mathrm{d}x=-\int_\Omega\frac{\partial f_k(x)}{\partial x_i}\varphi(x)\mathrm{d}x,$$

$$\int_\Omega f(x)\frac{\partial\varphi(x)}{\partial x_i}\mathrm{d}x=-\int_\Omega g_i(x)\varphi(x)\mathrm{d}x.$$

所以 $g_i(x)$也是 $f(x)$的弱广义导数.上面要注意，因为 $\varphi(x)\in C_0^1(\Omega)$而在$\partial\Omega$ 附近恒为 0，故$f(x)$，$g_i(x)$等仅仅是局部可积也不影响计算.

(弱)⇒(强).因为 $g_i(x)$是 $f(x)$之弱广义导数，故若令(2)中的 $\varphi(y)$为$\frac{1}{L\delta^n}j_\delta(x-y)$($x$ 当作参数)即有

$$M_\delta(g_i)(x)=\int_\Omega g_i(y)\varphi(x-y)\mathrm{d}y,$$

$$\frac{\partial}{\partial x_i}M_\delta(f)(x)=\int_\Omega f(y)\frac{\partial\varphi(x-y)}{\partial x_i}\mathrm{d}y=-\int_\Omega f(y)\frac{\partial\varphi(x-y)}{\partial y_i}\mathrm{d}y,$$

且

$$\frac{\partial}{\partial x_i}M_\delta(f)(x)=M_\delta(g_i)(x).$$

由前面讲的 $L^p_{\mathrm{loc}}(\Omega)$如何用磨光算子逼近，记 $M_\delta(f)(x)=f_\delta(x)\in L^p_{\mathrm{loc}}(\Omega)$，$M_\delta(g_i)(x)=g_{i\delta}(x)\in L^p_{\mathrm{loc}}(\Omega)$，即有 $\|f_\delta-f\|_{L^p(K)}\to 0$，$\|g_{i\delta}-g_i\|_{L^p(K)}\to 0$ ($K\subset\subset\Omega$ 是任意紧子集)，$\dfrac{\partial f_\delta}{\partial x_i}=g_{i\delta}$. 所以 $f$ 之强广义导数$\dfrac{\partial f}{\partial x_i}$即 $g_i(x)$，证毕.

这两部分证明的难度显然有区别. 前一部分我们将在以后看到，仅只使用了泛函的连续性，而这是一个很普遍的性质；后一部分则用了磨光技巧，这并不是处处都能使用的. 因此在许多数学问题中，一个概念常有强、弱两种推广，证明二者相同是不容易的事.

以下我们就只说广义导数而不再区分强弱.

**定义 4**　定义

$$W^{k,p}(\Omega)=\{f(x)\in L^p(\Omega),\text{且其广义导数 }\mathrm{D}^\alpha f\in L^p(\Omega),|\alpha|\leqslant k\}\tag{6}$$

并称它为 Sobolev 空间，我们也常用记号 $H^{k,p}(\Omega)$.

从强弱广义导数一致性的证明中我们看到 $W^{k,p}(\Omega)$之元必可通过磨光得到的 $C^\infty(\Omega)$函数在范数

$$\|f\|_{W^{k,p}}=\Big(\int_\Omega\sum_{|\alpha|\leqslant k}|\mathrm{D}^\alpha f|^p\mathrm{d}x\Big)^{1/p}\tag{7}$$

下求极限而得. 强广义导数定义中的 $f_k\in C^1(\Omega)$改成 $C^\infty(\Omega)$也是可以的(例如用 $f$ 之磨光 $M_{1/k}(f)(x)=f'_k$代替 $f_k$ 即可). 所以我们说 Sobolev 空间是 $C^\infty(\Omega)$经(7)完备化而得. 现在问，如果取一串 $f_m\in W^{k,p}$ 再作一次完备化，会不会得到更多新的东西？结论是否定的. 因为在 $\|_{Jm}\|_{W^{k,p}}$中取$\alpha=0$ 的一项可见 $\|_{Jm}-_{Jl}\|_{L^p}\leqslant\|_{Jm}-_{Jl}\|_{W^{k,p}}\to 0$，所以一定存在一个函数 $f\in L^p$ 使 $\|f_m-f\|_{L^p}\to 0$. 同理取$\left\{\dfrac{\partial f_m}{\partial x_i}\right\}$，又有$\left\|\dfrac{\partial f_m}{\partial x_i}-\dfrac{\partial f_l}{\partial x_i}\right\|_{L^p}\leqslant\|f_m-f_l\|_{W^{k,p}}\to 0$，所以$\dfrac{\partial f_m}{\partial x_i}$必有极限 $g_i\in L^p$. 由广义导数定义易知 $g_i=\dfrac{\partial f}{\partial x_i}$. 仿此类推可知$\{f_m\}$必有极限 $f\in W^{k,p}$. 这就是说

**定理 3**　$W^{k,p}(\Omega)=H^{k,p}(\Omega)$是一个完备空间.

如果用 $C_0^\infty(\Omega)$来作完备化，我们将记所得空间为 $H_0{}^{k,p}(\Omega)$或 $W_0{}^{k,p}(\Omega)$，它也是完备的. 但是其中之元并非具紧支集的函数. 若 $k\geqslant 1$，可以证明 $H_0{}^{k,p}(\Omega)$中之元在$\partial\Omega$ 上按一定意义为 0. 以上用到的完备空间与完备化的概念，将在第六章详细讨论.

**注 1**　在许多文献中，只在 $p=2$ 时才用记号 $H^{k,2}(\Omega)$或 $H_0{}^{k,2}(\Omega)$，但通常都将上标 2 略去.

**注 2**　我们也时常要用到 $W^{k,p}_{\mathrm{loc}}(\Omega)$或 $H^{k,p}_{\mathrm{loc}}(\Omega)$.

**2.**

了著名的狄拉克的 $\delta$ 函数. 这种对象称为广义函数，当代对广义函数的处理最常用的方法是利用对偶性. 下面我们就以此为线索讨论它.

第二章 §4 中已经指出，广义函数就是试验函数空间——例如上面讲的 $D(\Omega)$——上的连续线性泛函(但那里没有在“连续”二字上做文章). 有许多种不同的试验函数空间. 为简单起见，我们就考虑 $D(\Omega)$ 好了. 所谓泛函，就是一个线性映射：$l: D(\Omega)\to$ 常数(这个数可以是实数，可以是复数. 我们暂定为实数，于是泛函 $l$ 也就用欧氏配对 $\langle l,\varphi\rangle$ 表示. 在第五章中会看见此常数取为复数的情况，于是记号 $l(\varphi)$ 这时就表示埃尔米特配对 $(l,\varphi)$. 有的文献上用 $(\varphi, l)$，这不会造成大的困难). 所谓线性泛函，即对 $\varphi_1,\varphi_2\in D(\Omega)$，以及常数 $C_1, C_2$ 恒有

$$l(C_1\varphi_1+C_2\varphi_2)=C_1 l(\varphi_1)+C_2 l(\varphi_2).$$

所谓连续泛函，如果细讲就要涉及试验函数空间的拓扑结构问题，我们将在第六章中作初步讨论，现在就只粗略地说：如果取一串试验函数 $\{\varphi_k(x)\}$，而且设此序列在此试验函数空间中趋于 0，则要求 $l(\varphi_k)\to 0$. 这里有两个问题：一是什么叫在试验函数空间中趋于 0？我们要指出，这正是广义函数理论之所以有强大力量的关键所在，我们马上就要来讨论它. 二是一个数学对象的连续性是否可以用序列来刻画？这是拓扑学中的一个一般问题，我们将在第六章中稍微涉及.

**定义 5** $D(\Omega)$ 中试验函数序列 $\{\varphi_k(x)\}$ 在 $D(\Omega)$ 中趋于 0 (我们引入一个记号：$\varphi_k\to 0$ in $D(\Omega)$) 即指存在一共同的紧子集 $K\Subset\Omega$ 使所有 $\varphi_k(x)$ 均适合 $\operatorname{supp}\varphi_k\subset K$，而且对任意固定的重指标 $\alpha$，$\partial^\alpha\varphi_k(x)$ 在 $\Omega$ 上对 $x$ (而不必对 $\alpha$) 一致趋于 0.

$\operatorname{supp}\varphi_k\subset K$ 这个条件是重要的. 因为若不如此，可能 $\varphi_k$ 之极限函数并不是在 $\Omega$ 内有紧支集的函数，因而 $l(\varphi)$ 可能没有意义. 例如取 (4) 式定义的 $j_1(x)$ 为 $\varphi(x)$，则 $\varphi_k(x)=\dfrac{1}{k}\varphi(x)\to 0$ in $C(\mathbf{R}^n)$，但 $\dfrac{1}{k}\varphi\left(\dfrac{x}{k}\right)$ 则不是在 $D(\mathbf{R}^n)$ 中趋于 0，因为 $\operatorname{supp}\varphi\left(\dfrac{x}{k}\right)=\{x:|x|\leqslant k\}$ 不含于共同的紧支集中.

**定义 6** $D(\Omega)$ 上的连续线性泛函 $l$ 称为 $D'(\Omega)$ 广义函数(在不发生误解时就简称为广义函数)其集合记为 $D'(\Omega)$.

所谓连续性就指：当 $\varphi_k\to 0$ in $D(\Omega)$ 时，$l(\varphi_k)\to 0$.

下面看一些广义函数的例子.

**例 4** 设 $f(x)\in L^1_{\text{loc}}(\Omega)$，则对 $\varphi(x)\in D(\Omega)$，

$$l(\varphi)=\int_\Omega f(x)\varphi(x)\,\mathrm{d}x=\int_{\mathbf{R}^n} f(x)\varphi(x)\,\mathrm{d}x \tag{8}$$

是有意义的. 它自然是线性泛函. 又若 $\varphi_k(x)\to 0$ in $D(\Omega)$，则 $\varphi_k(x)$ 在 $\Omega$ 上对 $x$ 一致地趋于 0. 利用勒贝格控制收敛定理有 $l(\varphi_k)\to 0$，所以 (8) 定义了一个 $D'(\Omega)$ 广义函数. 粗略地说，一个局部可积函数就是一个广义函数，而且称为正则广义函数. 但还有不能由局部可积函数定义的广义函数，称为奇异广义函数，其中最著名的是狄拉克的 $\delta$ 函数.

**例 5** 设 $0\in\Omega$，对于 $\varphi\in D(\Omega)$，我们定义

$$\langle\delta,\varphi\rangle=\varphi(0). \tag{9}$$

它也是线性连续泛函，称为 $\delta$ 函数. 上面例 2 讲赫维赛德函数时就证明了它不可能是一个局部

可积函数. 那里尽管是对 $\Omega\subset\mathbf{R}^1$ 的情况来证明的，其实对高维的 $\mathbf{R}^n$ 也是成立的. 第二章 §4 中讲了，$\delta$ 函数其实是一个测度. 这不是偶然的. 由(9)我们可以看到

$$|\langle\delta,\varphi\rangle| = |\varphi(0)| \leqslant \sup_{\text{supp}\,\varphi}|\varphi(x)|.$$

$\delta$ 函数可以拓展到更大的空间 $C_0(\Omega)\supset D(\Omega)$ 上去. 在 $C_0(\Omega)$ 上我们用 $\sup|\varphi_k(x)|\to 0$ 以及 $\text{supp}\varphi_k\subset K$ 来定义 $\varphi_k(x)\to 0$ in $C_0(\Omega)$. 凡是适合这种不等式的 $D'(\Omega)$ 广义函数都是一种测度，称为 Radon 测度. 勒贝格测度也是一种 Radon 测度，这些我们都不能细讲了. 但我们看到了广义函数中不仅包含局部可积函数，而且还有很大一类测度. 但是还有更多的东西，例如

**例 6**　$f(x)=\dfrac{1}{x}$，$x\in\mathbf{R}^1$ 不是局部可积的，$\displaystyle\int_{-\infty}^{+\infty}\frac{1}{x}\varphi(x)\mathrm{d}x$，当 $\varphi(x)\in C_0^\infty(\mathbf{R}^1)$ 时，一般而言是没有意义的. 但是我们可以定义这个积分的哥西主值：

$$\text{v.p.}\quad\int_{-\infty}^{\infty}\frac{\varphi(x)}{x}\mathrm{d}x = \lim_{\varepsilon\to 0}\left[\int_{-\infty}^{-\varepsilon}+\int_{\varepsilon}^{+\infty}\right]\frac{\varphi(x)}{x}\mathrm{d}x. \tag{10}$$

注意到 $\varphi(x)\in C_0^\infty(\mathbf{R}^1)$，利用在 $x=0$ 处的泰勒公式

$$\varphi(x)=\varphi(0)+x\psi(x). \tag{11}$$

代入上式，并令 $\text{supp}\varphi\subset[-A,A]$，有

$$\begin{aligned}\left[\int_{-\infty}^{-\varepsilon}+\int_{\varepsilon}^{+\infty}\right]\frac{\varphi(x)}{x}\mathrm{d}x &= \left[\int_{-A}^{-\varepsilon}+\int_{\varepsilon}^{+A}\right]\frac{\varphi(0)}{x}\mathrm{d}x+\int_{A\geqslant|x|\geqslant\varepsilon}\psi(x)\mathrm{d}x\\ &= \int_{A\geqslant|x|\geqslant\varepsilon}\psi(x)\mathrm{d}x\end{aligned}$$

这里我们用到了 $\dfrac{\varphi(0)}{x}$ 对 $x$ 为奇函数因而第一项为 0. 代入(10)求极限，有

$$\text{v.p.}\quad\int_{-\infty}^{\infty}\frac{\varphi(x)}{x}\mathrm{d}x=\int_{-A}^{A}\psi(x)\mathrm{d}x. \tag{12}$$

它是线性泛函是明显的，利用拉格朗日公式可知 $\psi(x)=\varphi'(\theta x)$，$0\leqslant\theta\leqslant 1$，$x\in\text{supp}\,\varphi$. 因此，当 $\varphi_k(x)\to 0$ in $D(\mathbf{R}^1)$ 时，$\psi(x)$ 在 $\text{supp}\,\varphi$ 上一致收敛于 0，因而(12)趋向 0. 这样，由(10)所定义的线性泛函也是一个 $D'(\Omega)$ 广义函数，以下我们记作 $\text{v.p.}\dfrac{1}{x}$. 很容易想象到，它不是一个 Radon 测度，因为在证明其连续性时我们用到了试验函数的一阶导数. 但是下面我们要证明 $\text{v.p.}\dfrac{1}{x}=(\ln|x|)'$. 这里导数的意义将在下面解释.

在物理学特别是在量子力学中会遇到许多“奇异的”对象而时常都可以用广义函数去解释. 在数学中也有许多结果在用广义函数解释后更加明确. 本书后面经常要这样做，所以现在我们再讲一些广义函数的性质.

首先提到，广义函数由于只是泛函而不是本来意义下的函数，所以问它在某一点的值是多少是没有意义的. 我们时常也把例如 $\delta$ 函数写成 $\delta($

只不过是为了提醒一下，它的来源之一是局部可积函数乘上 $\varphi(x)$ 后再积分. 但是这并不是说广义函数没有局部性质. 例如 0 泛函作为一个广义函数 $f(x)$ 其定义就是 $f(x)$ 作用在一切试验函数 $\varphi(x)$ 上之值为 0，所以若 $\langle f,\varphi\rangle=0$ 对一切 $\varphi\in D(\Omega)$ 成立，就说 $f=0$ in $\Omega$，这是 $f=0$ 的整体定义. 它还有局部定义：我们说 $f$ 在 $x_0\in\Omega$ 之某个邻域 $U$ 中为 0，即指若 $\varphi\in D(\Omega)$ 之支集在 $U$ 中：$\operatorname{supp}\varphi\subset U$，均有 $\langle f,\varphi\rangle=0$. 现在证明

**定理 4**　若 $f(x)$ 在 $\Omega$ 之每一点的某邻域中均为 0，则 $f(x)$ 整体为 0，其逆亦真.

**证**　对每一点 $x\in\Omega$，如上所述都可以找到一个邻域 $U_x\subset\Omega$. 对任一 $\varphi(x)\in D$，任取一点 $x\in\operatorname{supp}\varphi$，必有其相应的邻域 $U_x$，很明显这些邻域构成 $\operatorname{supp}\varphi$ 的开覆盖，而由有限覆盖定理，它必有一个有限子覆盖：$\{U_1,\cdots,U_N\}$. 作从属于它的单位分解

$$1=\sum_{i=1}^{N}\chi_i(x),\quad \chi_i(x)\in C_0^\infty(\Omega),\operatorname{supp}\chi_i\subset U_i,$$

于是令 $\varphi_i(x)=\chi_i(x)\varphi(x)$，其支集在 $\chi_i$ 的邻域 $U_i$ 中. 但由假设，$f$ 在 $U_i$ 上为 0，故

$$\langle f,\varphi\rangle=\sum_{i=1}^{N}\langle f,\varphi_i\rangle\equiv 0,$$

即 $f(x)$ 整体为 0.

逆定理就不必证明了.

**定义 7**　广义函数 $f(x)$ 之支集 $\operatorname{supp} f$ 定义为

$$\operatorname{supp} f=\{x:\quad x\text{ 有一个邻域使在其上 }f\neq 0\}.$$

$\operatorname{supp} f$ 显然是闭集.

与此类似还可以考虑广义函数的光滑性. 显然光滑函数都是局部可积的. 若在 $x\in\Omega$ 之某一邻域 $U$ 中，广义函数 $f(x)$ 与某一 $C^\infty(U)$ 函数 $g(x)$ 相等：即 $f(x)-g(x)$ 作为广义函数局部为 0，我们就说 $f(x)$ 在 $x$ 附近为 $C^\infty$. 若 $f(x)$ 在某点附近不是 $C^\infty$ 的，就说 $x$ 点是 $f(x)$ 之奇点.

**定义 8**　$f(x)\in D'(\Omega)$ 之奇支集即其奇点之集合：

$$\operatorname{sing\,supp} f=\{x:\quad f(x)\text{ 在其任一邻域中均不为 }C^\infty\}.$$

奇支集 $\operatorname{sing\,supp} f$ 显然也是闭集.

**3. 广义函数的代数性质**　广义函数之空间 $D'(\Omega)$ 是线性空间这是显然的. 现在要问，若“自变量” $x$（上面已经说了，$x$ 其实是试验函数的自变量）经过线性变换，$D'(\Omega)$ 之元相应应如何变化？由于广义函数作为试验函数空间上的连续线性泛函最简单的形式是积分，所以这里讲的自变量的变换应该推广积分中的积分变量的线性变换（一般的变量变换涉及很困难的问题），先不妨设 $f(x)$ 是正则广义函数，$\varphi(x)\in D(\Omega)$，则若 $A:\mathbf{R}^n\to\mathbf{R}^n$ 是一个非奇异线性变换，或者说 $A$ 是一个 $n\times n$ 矩阵且 $\det A\neq 0$，$a\in\mathbf{R}^n$，则对 $y\in\mathbf{R}^n$

$$\langle f(Ay-a),\varphi(y)\rangle=\int_{\mathbf{R}^n} f(Ay-a)\varphi(y)\,\mathrm{d}y.$$

令 $x=Ay-a$，或 $y=A^{-1}(x+a)$，代入上式有

$$\begin{aligned}\langle f(Ay-a),\varphi(y)\rangle&=\frac{1}{|\det A|}\int_{\mathbf{R}^n} f(x)\varphi(A^{-1}(x+a))\,\mathrm{d}x\\&=\frac{1}{|\det A|}\langle f(x),\varphi(A^{-1}(x+a))\rangle.\end{aligned}\tag{13}$$

这里自然仍有 $\varphi(A^{-1}(x+a))=\psi(x)\in D(\Omega_x)$，$\Omega_x$ 是 $\mathbf{R}_y^n$ 中的 $\Omega$ 在映射 $x=Ay-a$ 下的像.这里我们要注意开集与紧集在非奇异线性变换下仍变为开集与紧集，这样才知道 supp $\psi(x)$ 是开集 $\Omega_x$ 的紧子集.

对于一般的奇异的广义函数 $f(x)$，(13)式中间一项积分没有意义了，但两端两项仍有意义，于是我们给出

**定义和定理 9**　设 $f(x)\in D'(\Omega_x)$，$A:\mathbf{R}^n\to\mathbf{R}^n$ 是非奇异线性变换，$a\in\mathbf{R}^n$，则定义 $f$ 在变换 $x=Ay-a$ 下的像 $f(Ay-a)\in D'(\Omega_y)$ 为

$$\langle f(Ay-a),\varphi(y)\rangle=\frac{1}{|\det A|}\langle f(x),\varphi(A^{-1}(x+a))\rangle. \tag{14}$$

我们用“定义和定理”这个说法是因为 $f(Ay-a)\in D'(\Omega)$ 有待证明.但是上式右方已经是一个连续线性泛函，这一点很容易看出.所以我们已经证明了这一点.

现在看定义和定理 9 的一些特例.

如果 $A=I$，就得到平移 $x=y-a$.就是说 $y$ 坐标系是由 $x$ 坐标系向左移动 $a$ 而来.这时，$\det A=1$，故(14)成为

$$\langle f(y-a),\varphi(y)\rangle=\langle f(x),\varphi(x+a)\rangle.$$

$\varphi(x+a)$ 的图像可由 $\varphi(x)$ 的图像向左移动 $a$ 而得，所以，如果想把 $f(y)$ 向右移一个 $a$ 而得新的泛函 $f(y-a)$，只需将被它作用于其上的试验函数 $\varphi(x)$ 图像向左移动 $a$ 就行了.例如

$$\langle\delta(y-a),\varphi(y)\rangle=\langle\delta(x),\varphi(x+a)\rangle=\varphi(a).$$

从所附的图 4-6-2 可以形象地看到这件事，而且也可以形象地理解“对偶”是什么意思.

其次看 $A=cI$，$a=0$，这里 $c>0$，即线性变换是沿向量 $x$ 方向放大或缩小 $c$ 倍，这时 $\det A=c^n$，于是(14)给出

$$\langle f(cy),\varphi(y)\rangle=c^{-n}\langle f(x),\varphi(x/c)\rangle. \tag{15}$$

图 4-6-2

研究函数性质时，齐性函数是十分有趣——且特别有用的.$\lambda$ 次齐性函数即适合关系式

$$f(cx)=c^{\lambda}f(x),\qquad c>0 \tag{16}$$

的函数.这里与通常微积分教本不同在于我们限制了 $c>0$.这是因为上式从几何上讲的是如果自变量在 $Ox$ 半射线方向放大 $c$ 倍,函数值 $y=f(x)$ 将在半射线 $Oy$ 方向上放大 $c^{\lambda}$ 倍,若 $c<0$ 则自变量将在 $Ox$ 的反方向上放大 $c$ 倍,$y$ 放大的倍数 $c^{\lambda}$ 可能是复数.更重要的是,我们常要在 $x$ 空间以原点为顶点的锥内讨论一个数学或物理问题,沿 $Ox$ 的反方向放大则会越出锥外,这时常在物理上很难理解.若附加上 $c>0$ 的条件,齐性函数也时常称为正齐性函数,但这并不是说齐性函数限于取正值.

若一个 $D'(\Omega)$ 广义函数 $f(x)$ 适合(16)式,这里 $c>0$,我们也说它是一个 $\lambda$ 次正齐性 $D'(\Omega)$ 广义函数.这时,我们有,对于 $\varphi\in D(\Omega)$

$$\langle f(cy),\varphi(y)\rangle=c^{-n}\left\langle f(x),\varphi\left(\frac{x}{c}\right)\right\rangle=c^{\lambda}\langle f(y),\varphi(y)\rangle. \tag{17}$$

因此

$$\langle f,\varphi\rangle=c^{-(\lambda+n)}\left\langle f(x),\varphi\left(\frac{x}{c}\right)\right\rangle. \tag{18}$$

特别是,若 $f(y)=\delta(y)$,由(17) 有

$$\langle\delta(cy),\varphi(y)\rangle=c^{-n}\left\langle\delta(x),\varphi\left(\frac{x}{c}\right)\right\rangle=c^{-n}\varphi(0)=c^{-n}\langle\delta,\varphi\rangle,$$

因此

$$\delta(cy)=c^{-n}\delta(y),$$

即是说 $\delta$ 函数是 $-n$ 次正齐性 $D'(\Omega)$ 广义函数.

我们知道,对 $\lambda$ 次正齐性可微函数 $f(x)$,有欧拉公式成立:

$$\sum_{i=1}^{n}x_i\frac{\partial f}{\partial x_i}(x)=\lambda f(x). \tag{19}$$

对 $\lambda$ 次正齐性广义函数它也是对的.这与(18)相比较有一点怪:在(18)中齐性指数是 $c+n$,$n$ 的来源自然是“积分的变量变换”.可是到了(19)式,这个 $n$——积分区域的维数——又不见了,其原因下面讲到广义函数的求导时就自然明白了.

若 $f(x)\in D'(\Omega)$ 在线性变换下不变,就称之为此变换下不变的广义函数.例如若 $f(x)\in D'(\Omega)$ 相对于某个固定的平移 $a$ 是不变的,就称之为以 $a$ 为周期的广义函数:

$$\langle f(x-a),\varphi(x)\rangle=\langle f(y),\varphi(y+a)\rangle=\langle f(y),\varphi(y)\rangle. \tag{20}$$

当然表面上看这似乎意味着 $\varphi(y+a)=\varphi(y)$,但其实不是.因为(20)中的 $\varphi\in D(\Omega)$ 应该是任意试验函数而不一定有周期(再说,一个周期函数除非恒等于0,一定不会是紧支集的),所以问题出在 $f(x)$.我们可以举一个周期广义函数之例:$f(x)=e^{i\omega x}$,它不是可积的,但是是局部可积的($f(x)\equiv1$ 相应于 $\omega=0$ 也只是局部可积的),它以 $a=2\pi n/\omega$, $n=0,\pm1,\pm2,\cdots$ 为周期(或者说以 $\frac{2\pi}{\omega}$ 为周期).这是一个很有用的广义函数.

若 $A=-I$, $a=0$,则线性变换为 $x=Ay-a=-y$,即一个反射.在反射下不变的广义函数

$$f(-y)=f(y)$$

称为偶广义函数;若 $f(-y)=-f(y)$ 则称为奇广义函数.

**4. 广义函数的分析运算** 因为广义函数建立后最直接的应用是在线性偏微分方程上,所以我们首先要考察一个线性偏微分算子

$$P(x,\partial_x)=\sum_{|\partial|\leqslant m}a_\alpha(x)\partial_x^\alpha$$

如何作用于一个广义函数上.这里包括了两个运算,一是乘以系数,它们是函数 $a_\alpha(x)$;二是求导.我们依次来讨论.

首先讨论函数 $a(x)$与 $f(x)\in D'(\Omega)$之乘积,仿照 $f(x)$局部可积时广义函数的定义

$$\int[a(x)f(x)]\varphi(x)\mathrm{d}x=\int f(x)[a(x)\varphi(x)]\mathrm{d}x,$$

我们应定义 $a(x)f(x)$如下:

$$\langle a(x)f(x),\varphi(x)\rangle=\langle f(x),a(x)\varphi(x)\rangle.\tag{21}$$

但为此就需要 $a(x)\varphi(x)\in D(\Omega)$.这就需要 $a(x)\in C^\infty(\Omega)$,还需要当 $\varphi_k(x)\to 0$ in $D(\Omega)$时 $a(x)\varphi_k(x)\to 0$ in $D(\Omega)$.这也是容易的,因为 $\mathrm{supp}\varphi_k$ 均含于$\Omega$之一个共同紧子集$K$中,所以 $\mathrm{supp}\ a(x)\varphi_k(x)\subset K$;又因 $\varphi_k$ 的任一固定阶导数一致收敛于 0,$a\varphi_k$ 的同一阶导数当然也一致收敛于 0,总之 $a(x)\varphi_k(x)\to 0$ in $D(\Omega)$,因此

$$\lim_{k\to\infty}\langle a(x)f(x),\varphi_k(x)\rangle=\lim_{k\to\infty}\langle f(x),a(x)\varphi_k(x)\rangle=0.$$

所以 $af\in D'(\Omega)$.总之我们看到,不是任意函数都可以与 $f\in D'(\Omega)$相乘,我们需要 $a(x)$: $D'(\Omega)\to D'(\Omega)$(下面还要证明这是一个连续映射).适合这种要求的函数称为 $D'(\Omega)$乘子.上面我们看到 $C^\infty(\Omega)$函数都是 $D'(\Omega)$乘子,也易见,若 $a(x)\notin C^\infty(\Omega)$,$af$ 就可能不属于$D(\Omega)$.

读者自然会想到:两个广义函数可否相乘?一般说来这是不可能的.问题在于我们不能随心所欲地胡乱定义某个数学概念.我们想要定义的广义函数之乘积既是通常函数乘积之推广,则它自然应该具备通常函数乘积的某些基本性质.这里我们希望能保持的是乘法的一些基本运算律,而我们会发现恰好是乘法的结合律$(AB)C=A(BC)$现在不可能保持.

如果乘法要有意义,我们当然希望$1\cdot f(x)=f(x)$,$f\in D'(\Omega)$,因为 1 是 $D'(\Omega)$乘子;也希望$0\cdot f(x)=0$,因为 0 也是 $D'(\Omega)$乘子.同时注意 $x\delta(x)=0$ $(x\in\mathbf{R}^1)$,这是因为 $x$ 也是$D'(\mathbf{R}^1)$乘子,故对 $\varphi(x)\in D(\mathbf{R}^1)$应有

$$\langle x\delta(x),\varphi(x)\rangle=\langle\delta(x),x\varphi(x)\rangle=[x\varphi(x)]_{x=0}=0.$$

但这样一来,$\left(\frac{1}{x}\cdot x\right)\delta(x)=1\cdot\delta(x)=\delta(x)$,但$\frac{1}{x}\cdot(x\delta(x))=0$(上面$\frac{1}{x}$指 v. p. $\frac{1}{x}$,$\frac{1}{x}\cdot x=1$ 应该证明,但是很容易证,故略去).这样一来,如果我们勉强去定义广义函数之乘积,则连乘子运算也不能包括在内.近来的研究表明,广义函数相乘的“障碍”在于它的奇性.时常有人想求 $\delta(x)$的“平方”,按现在的知识水平,还没有一种公认的处理方法.

其次要考虑广义函数的导数,这时我们仍然模仿分部积分法而给出如下的

**定义和定理 10**

到的，为了证明其连续性，我们只限于(23)式，并以 $\varphi_k\in D(\Omega)$代替其中的 $\varphi$，而来证明当 $\varphi_k\to 0$ in $D(\Omega)$时，(23)左方之式也趋于 0. 这是很容易的. 记 $\psi_k^\alpha=\dfrac{\partial^\alpha\varphi_k}{\partial x^\alpha}$，则易见 $\psi_k^\alpha\in D(\Omega)$，而且 supp $\psi_k^\alpha\subset$supp $\varphi_k\subset K$($\Omega$ 的一个公共的紧子集)，$\psi_k^\alpha$ 的某一固定阶导数$\partial^\beta\psi_k^\alpha=\dfrac{\partial^{\alpha+\beta}\varphi_k}{\partial x^{\alpha+\beta}}$也是 $\varphi_k$ 的某个固定阶导数，故当 $k\to\infty$时对 $x$ 一致趋于 0，因此(23)的式右由于 $f\in D'(\Omega)$应趋于 0，从而式左亦然. 这还告诉了我们，$D'(\Omega)$广义函数可以求任意阶导数.

下面讲一些有关导数的性质，有些很简单的就不证明了. 首先讨论原函数问题. 为此，我们限令 $\Omega\subset\mathbf{R}^1$. 若 $f\in D'(\Omega)$，$g\in D'(\Omega)$，而且$\dfrac{\mathrm{d}f}{\mathrm{d}x}=g$ 就说$f$ 是$g$ 的原函数(其实只是广义函数). 对于古典意义的函数，并非一切 $g(x)$均有原函数，甚至黎曼可积函数也不一定有原函数. 所以 §1 的达布定理只说，当 $g(x)$既是黎曼可积又有原函数存在(若 $g(x)$连续，则两个条件都满足)，牛顿－莱布尼茨公式才成立，但是广义函数给了我们一个惊喜，因为我们有

**定理 5** 每一个 $D'(\mathbf{R}^1)$广义函数均有无穷多个原函数存在，且任意两个原函数只相差一个常值函数；反之由任一原函数加上常值函数即可得出全部原函数.

**证** 证明分两部分. 首先设 $g(x)\in D'(\mathbf{R}^1)$确有原函数 $f(x)\in D'(\mathbf{R}^1)$存在：$\dfrac{\mathrm{d}f}{\mathrm{d}x}=g$，我们要求出 $f(x)$的表达式. 其次再证明这个表达式确实定义了一个 $D'(\mathbf{R}^1)$广义函数，而且确是 $g(x)$的原函数. 由此，原函数是存在的.

先看第一部分，要找 $D'(\mathbf{R}^1)$中一个 $f(x)$，只需知道它作用在任意 $\varphi(x)\in D(\mathbf{R}^1)$上之值即可. 若 $\varphi(x)$有原函数 $\psi(x)$(这是不成问题的)在 $D(\mathbf{R}^1)$中(这一点却不一定)：$\varphi(x)=\dfrac{\mathrm{d}\psi}{\mathrm{d}x}$，$\psi\in D(\mathbf{R}^1)$，则 $\langle f,\varphi\rangle=\langle f,\psi'\rangle=-\langle f',\psi\rangle=-\langle g,\psi\rangle$. 所以，$f(x)$作用在这种 $\varphi(x)$上所得的值是清楚的，那么，什么样的 $\varphi(x)$才有 $D(\mathbf{R}^1)$原函数呢？这种原函数如果存在必是 $\psi(x)=\int_{-\infty}^x\varphi(t)\mathrm{d}t$. 因为已设 supp $\varphi\subset[-M,M]$，至少当 $x\leqslant -M$ 时$\psi(x)=0$，但我们还需要当 $x$ 充分大时$\psi(x)=0$，所以应要求$\int_{-\infty}^{+\infty}\varphi(t)\mathrm{d}t=0$. 说来也巧，如果 $\varphi(x)$适合这个条件，则当 $x>M$ 时，$\psi(x)=\int_{-\infty}^x\varphi(t)\mathrm{d}t=\int_{-\infty}^x\varphi(t)\mathrm{d}t+\int_x^{+\infty}\varphi(t)\mathrm{d}t=\int_{-\infty}^{\infty}\varphi(t)\mathrm{d}t=0$. 这就是我们要求于$\varphi(x)$的性质. 因此，我们记

$$H=\left\{\varphi(x)\in D(\mathbf{R}^1),\quad \int_{-\infty}^{+\infty}\varphi(x)\mathrm{d}x=0\right\}. \tag{24}$$

$H$ 当然只是$D(\mathbf{R}^1)$真子空间，问题在于这个 $\varphi(x)$是否在 $H$ 中. $\varphi\in C^\infty$不成问题，但是 $D(\mathbf{R}^1)$中的元必可分为两项之和，一项在 $H$ 中，另一项是 $\varphi_0(x)$的倍数，这里的 $\varphi_0(x)\in D(\mathbf{R}^1)$，而$\int_{-\infty}^{+\infty}\varphi_0(t)\mathrm{d}t=1$. 例如令

$$\varphi_0(x)=\begin{cases}\alpha\exp[1/(x^2-1)], & x\leqslant 1,\\ 0 & x\geqslant 1,\end{cases}$$

而 $\alpha$ 适当选取即得. 事实上,若 $\varphi(x)\in D(\mathbf{R}^1)$,令 $A=\int_{-\infty}^{\infty}\varphi(t)\mathrm{d}t$,则显然

$$\varphi_H(x)=\varphi(x)-A\varphi_0(x)\in H,$$

或者说,$D(\mathbf{R}^1)$中任一元均可分解如下:

$$\varphi(x)=\varphi_H(x)+A\varphi_0(x). \tag{25}$$

$\varphi_H(x)$之 $D(\mathbf{R}^1)$原函数是

$$\psi(x)=\int_{-\infty}^{x}\varphi_H(t)\mathrm{d}t=\int_{-\infty}^{x}\varphi(t)\mathrm{d}t-A\int_{-\infty}^{x}\varphi_0(t)\mathrm{d}t. \tag{26}$$

现在用 $f(x)$作用于(25)两端得到

$$\begin{aligned}\langle f,\varphi\rangle&=\langle f,\frac{\mathrm{d}\psi}{\mathrm{d}x}\rangle+A\langle f,\varphi_0\rangle\\&=-\langle g,\psi\rangle+C\int_{-\infty}^{+\infty}\varphi(t)\mathrm{d}t.\end{aligned} \tag{27}$$

第一部分证毕,再看第二部分,就用(27)式来定义 $f$,它是线性泛函是显然的,为了看它的连续性,取一串 $\varphi_k(x)\to0$ in $D(\mathbf{R}^1)$, $\operatorname{supp}\varphi_k\subset[-M,M]$,用 $\varphi_k$ 代替(26)和(27)中的 $\varphi$,于是 $\psi$ 也变成相应的 $\psi_k(x)$. 由(26)第一个等号易见: $\operatorname{supp}\psi_k\subset[-M_1,M_1]$,这里 $M_1=\max(M,1)$,而且当 $k\to+\infty$时其各阶导数均对 $x\in\mathbf{R}^1$ 一致收敛于 0,即是说 $\psi_k(x)\to0$ in $D(\mathbf{R}^1)$. 由(27)即知 $\langle f,\varphi_k\rangle\to0$. 所以,(27)确实定义了一个 $D'(\mathbf{R}^1)$广义函数. 我们只需证明$\frac{\mathrm{d}f}{\mathrm{d}x}=g$ 即可,为此,用 $\varphi'(x)$代替(26)中的 $\varphi$,则 $A=\int_{-\infty}^{+\infty}\varphi'(t)\mathrm{d}t=\varphi(x)\Big|_{-\infty}^{+\infty}=0$,而(26)给出 $\psi(x)=\int_{-\infty}^{x}\varphi'(t)\mathrm{d}t=\varphi(x)$,因此

$$\langle f,\varphi'\rangle=-\langle g,\varphi\rangle.$$

而 $f$ 确是 $g$ 的原函数. 证毕.

**推论 6** 若 $f(x)\in D'(\mathbf{R}^1)$且$\frac{\mathrm{d}f}{\mathrm{d}x}=0$,则 $f(x)\equiv\mathrm{const}$.

注 我们是对 $D'(\mathbf{R}^1)$证明了这个定理的. 但实际上它对 $D'(\Omega)$也成立,这里 $\Omega\subset\mathbf{R}^n$ 是一开集,但定理的结论中"只相差一个常值函数"要改成"只相差一个局部常值函数",即在 $\Omega$ 的不同连通分支中可能取不同的常数值的函数. 这里强调一下连通分支是必要的,例如我们在第二章§4 中就已看到$\frac{\mathrm{d}H(x)}{\mathrm{d}x}=\delta(x)$, $H(x)$是赫维赛德函数

$$H(x)=\quad 1,\quad x\geqslant$$

果把这两个连通分支合起来，这时得到 $\Omega=\mathbf{R}^1\setminus\{0\}$，我们只会得到

$$\frac{\mathrm{d}f}{\mathrm{d}x}=(C_+-C_-)\delta(x).$$

$C_+-C_-$ 正是 $\Omega$ 上的函数 $f(x)$ 在 $x=0$ 处的跃度.

这件事绝非偶然，在许多物理问题中，例如在量子力学的散射问题中，我们常会遇到在两个不同区域中研究某一个物理量，而在两个区域的连接处给出一定的"边值条件"把它们连接起来. 在每个区域中我们实际上并不在 $D'(\Omega)$ 中考虑问题而是在通常的比较光滑的函数之空间中考虑问题，例如在古典意义下取导数. 如果不把这两个区域分开考虑，而从整体上的 $\mathbf{R}^1$ 上来看问题，我们应把这些函数都看作 $D'(\mathbf{R}^1)$ 之元，而在广义函数意义下求导数，这样就会出现 $\delta(x)$. 因此，问题的实质在于我们究竟采取哪一种观点：$D'(\mathbf{R}^1)$ 还是 $x\gtrless 0$ 处的光滑函数以及连接二者的边值条件. 两者实质是相通的. 由此也就提出了一个问题：设有 $\mathbf{R}^1$ 上的函数 $f(x)$，它在 $x>0$ 和 $x<0$ 处均有一阶连续导数(这是古典意义的导数，我们记作 $f'_{\mathrm{cl}}$.) $f(x)$ 在 $x=0$ 处又有第一类间断点，记其跃度为 $[f]_0=f(0+)-f(0-)$，我们仍记 $f$ 之广义函数导数为 $f'$，现问 $f'$，$[f]_0$ 与 $f'_{\mathrm{cl}}$ 三者关系如何？有

**定理 7**　对上述 $f(x)$，我们有

$$f'(x)=f'_{\mathrm{cl}}(x)+[f]_0\delta(x). \tag{28}$$

**证**　考虑函数

$$F(x)=f(x)-[f]_0H(x).$$

显然 $F(x)\in C^1(\mathbf{R}^1\setminus 0)$，因而其古典意义下的导数与广义函数导数一致，但当 $x\neq 0$ 时，$F'_{\mathrm{cl}}=f'_{\mathrm{cl}}$，所以

$$F'(x)=f'(x)-[f]_0\delta(x),\ F'_{\mathrm{cl}}(x)=f'_{\mathrm{cl}}(x),$$

因此当 $x\neq 0$ 时

$$f'_{\mathrm{cl}}(x)=f'-[f]_0\delta(x).$$

此即(28)式. 不过这里的 $f'_{\mathrm{cl}}(x)$ 理解为古典意义下的导数所定义的广义函数. 所以我们不加注 $x\neq 0$. 再则，$\delta(x)$ 在某点之值是没有意义的，而应看作一个整体，所以也没有必要再注明 $x\neq 0$.

这是一个非常有用的结论，因为它把"集中分布"(以 $\delta(x)$ 为代表)与"边值条件"连接起来了. 我们只在一维情况讨论了它，其实在高维情况下也有类似结论.

古典意义下导数的性质有许多在 $D'(\Omega)$ 理论中仍成立. 例如

(1) 莱布尼茨公式. 设 $a(x)$ 为 $D'(\Omega)$ 乘子，$f(x)\in D'(x)$ 则

$$\partial[a(x)f(x)]=a(x)\partial f(x)+[\partial a(x)]f(x).$$

证略.

(2)　$$\frac{\partial^2 f}{\partial x_1\partial x_2}=\frac{\partial^2 f}{\partial x_2\partial x_1}.$$

证明也略去. 在通常微积分教本中会介绍一些函数 $f$ 使 $\dfrac{\partial^2 f(0)}{\partial x_1\partial x_2}\neq\dfrac{\partial^2 f(0)}{\partial x_2\partial x_1}$. 但时常只在个别点或个别曲线、曲面上如此. 在广义函数中，则高阶偏导数作为一个广义函数整体与求导次序无关.

(3) 齐性广义函数的欧拉公式. 设 $f(x)\in D'(\mathbf{R}^n)$ 为 $\lambda$ 次正齐性广义函数，则以下的欧拉公式成立

$$\sum_{i=1}^{n} x_i \frac{\partial f(x)}{\partial x_i} = \lambda f(x), \tag{29}$$

事实上,任取一个 $\varphi(x) \in D(\mathbf{R}^n)$,我们有

$$\langle \sum_{i=1}^{n} x_i \frac{\partial f}{\partial x_i}, \varphi \rangle = -\sum_{i=1}^{n} \langle f, \frac{\partial (x_i \varphi)}{\partial x_i} \rangle = -n\langle f, \varphi \rangle - \sum_{i=1}^{n} \langle f, x_i \frac{\partial \varphi}{\partial x_i} \rangle. \tag{*}$$

另一方面,由 $\lambda$ 次正齐性广义函数之定义

$$\langle f, \varphi \rangle = c^{-(\lambda+n)} \langle f(x), \varphi(\frac{x}{c}) \rangle, \quad c > 0.$$

双方对 $c$ 求导,再令 $c=1$,有

$$0 = -(\lambda+n)\langle f, \varphi \rangle - \sum_{i=1}^{n} \langle f, x_i \frac{\partial \varphi}{\partial x_i} \rangle,$$

即

$$-n\langle f, \varphi \rangle - \sum_{i=1}^{n} \langle f, x_i \frac{\partial \varphi}{\partial x_i} \rangle = \lambda \langle f, \varphi \rangle.$$

代入(*)式即得(29)式,我们现在看到,由“积分的变量变换”而多出来的 $n$ 是怎样消除了的.

**5. 广义函数序列的收敛** 第二章§4末尾,曾引用了狄拉克的话说 $\delta$ 函数并不是通常意义的函数,而是它们的极限,并且引用了一个例子,用这种思想证明了数学分析的一个重要结论:魏尔斯特拉斯定理.下面我们引入

**定义 11** 设 $\{f_k\} \subset D'(\Omega)$, $f \in D'(\Omega)$,使得对于一切 $\varphi(x) \in D(\Omega)$ 均有

$$\lim_{k\to\infty} \langle f_k, \varphi \rangle = \langle f, \varphi \rangle, \tag{30}$$

就说 $f_k \to f$ in $D'(\Omega)$.

注意,这个定义不只是说 $\lim\limits_{k\to\infty} \langle f_k, \varphi \rangle$ 存在,还要求极限是 $D(\Omega)$ 上一个连续线性泛函.因为我们暂时无法肯定这个极限也自动地定义 $D(\Omega)$ 上的一连续线性泛函,所以在定义中事先明确规定了 $f \in D'(\Omega)$.我们可以证明这个极限确实自动定义了一个 $D'(\Omega)$ 广义函数,但是我们不去证明这一点,因为这涉及了 $D'(\Omega)$ 作为 $D(\Omega)$ 的对偶空间有什么样的拓扑结构的问题.由于它过于专门,我们只好略去不论,而把极限为一连续线性泛函这一事实归入定义11中.我们只说一下,在文献中这个定义中的收敛称为弱收敛,它是非常广泛的.例如:设 $\{f_k(x)\}$ 是正则广义函数,即每个 $f_k(x)$ 都是 $\Omega$ 上局部可积函数,若此序列在 $\Omega$ 的每个紧子集 $K \subset \Omega$ 均一致收敛于 $f(x)$, $f(x)$ 也一定是 $\Omega$ 上的局部可积函数,而且易证 $\lim\limits_{k\to\infty} f_k(x) = f(x)$ in $D'(\Omega)$.然而它有许多古典的一致收敛性所没有的性质,例如

**定理 8** 若 $\lim\limits_{k\to\infty} f_k = f$ in $D'(\Omega)$,则对任意重指标 $\alpha$,有 $\lim\limits_{k\to+\infty} \partial^\alpha f_k = \partial^\alpha f$.

请读者自行证明.

这个收敛性之所以重要在于每个 $D'(\Omega)$ 广义函数都可以表示成为“很好的”函数(甚至 $D(\Omega)$ 函数)之 $D'(\Omega)$ 极限.在物理学上特别重

$$I_k = \left[\int_{-1}^{1} (1-x^2)^k \mathrm{d}x\right]^{-1},$$

是一个 $\delta$ 序列，而且我们常称之为朗道序列(这里的朗道是指德国数学家 E. Landau，而不是前苏联物理学家 L. Landau). 上一节中我们又指出正态分布的高斯核 $\frac{1}{2\sqrt{\pi Dt}}\exp[-x^2/4Dt]$，$t>0$ 也是一个 $\delta$ 序列. 不过这里没有指标 $k$，而以连续变化的 $1/t$ ($t\to 0+$ 时 $1/t\to+\infty$) 来代替 $k$. 现在我们来证明

**定理 9**(高斯序列) 当 $t\to 0+$ 时，$q(x,t)=\frac{1}{2\sqrt{\pi Dt}}\exp[-x^2/4Dt]\to\delta(x)$ in $D'(\mathbf{R}^1)$.

**证** 任给 $\varphi(x)\in D(\mathbf{R}^1)$. 很明显，因为

$$\frac{1}{2\sqrt{\pi Dt}}\int_{-\infty}^{\infty}\exp[-x^2/4Dt]\mathrm{d}x = 1,$$

所以

$$\varphi(0)=\frac{1}{2\sqrt{\pi Dt}}\int_{-\infty}^{\infty}\varphi(0)\exp[-x^2/4Dt]\mathrm{d}x .$$

于是我们有

$$\begin{aligned}&\langle q(x,t),\varphi(x)\rangle-\varphi(0)\\ &=\frac{1}{2\sqrt{\pi Dt}}\int_{-\infty}^{\infty}[\varphi(x)-\varphi(0)]\exp[-x^2/4Dt]\mathrm{d}x.\end{aligned}$$

把积分分为两部分，一部分在 $[-\delta,\delta]$ 上记作 $I_1$，其余的是在 $\mathbf{R}^1\setminus[-\delta,\delta]$ 上的积分记作 $I_2$. 任给一个 $\varepsilon>0$，因为 $\varphi(x)$ 在 $[-\delta,\delta]$ 中是一致连续的，故当 $\delta$ 充分小时有 $|\varphi(x)-\varphi(0)|<\frac{\varepsilon}{2}$，$|x|\leqslant\delta$. 因此

$$\begin{aligned}|I_1|&\leqslant\frac{1}{2\sqrt{\pi Dt}}\int_{-\delta}^{\delta}|\varphi(x)-\varphi(0)|\exp[-x^2/4Dt]\mathrm{d}x\\ &\leqslant\frac{\varepsilon}{2}\cdot\frac{1}{2\sqrt{\pi Dt}}\int_{-\infty}^{\infty}\exp[-x^2/4Dt]\mathrm{d}x=\frac{\varepsilon}{2}.\end{aligned}$$

现在固定 $\delta$ 并令 $t\to 0$. 因为 $\varphi(x)\in D(\mathbf{R}^1)$，故有最大值 $M$，而在 $I_2$ 中 $|\varphi(x)-\varphi(0)|\leqslant 2M$. 在 $I_2$ 中 $|x|\geqslant\delta$，从而 $\exp[-x^2/4Dt]\leqslant\exp[-\delta^2/8Dt]\exp[-x^2/8Dt]$，因此

$$|I_2|\leqslant\frac{\exp[-\delta^2/8Dt]}{2\sqrt{\pi}}$$

而定理证毕.

上一节中我们说过 $q(x,t)$ 是扩散方程

$$\frac{1}{D}\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2},\ -\infty<x<+\infty,\ t>0$$

的基本解.这就是说,当 $t>0$ 时 $q(x,t)$ 适合上面的偏微分方程(这一点读者很容易自己证),而当 $t\to 0$ 时,$q(x,t)\to\delta(x)$ in $D'(\Omega)$.这就是我们上面证明的事实.基本解是偏微分方程理论的基本工具.有了广义函数理论就能给出系统的基本解理论.这是广义函数论之所以重要的一个很大的理由.

我们不妨把这里的证明与第二章§4关于魏尔斯特拉斯定理的证明比较一下,立刻可以见到它们是十分相似的.实际上,我们常用的 $\delta$ 序列常是这样的光滑函数序列 $\{K_n(x)\}$:

(i) $K_n(x)\geqslant 0$;

(ii) $\int_{-\infty}^{\infty}K_n(x)\mathrm{d}x=1$;

(iii) 不论如何取 $\delta>0$, $\int\limits_{|x|\geqslant\delta}K_n(x)\mathrm{d}x\to 0\quad(n\to\infty)$,而 $\int\limits_{|x|<\delta}K_n(x)\mathrm{d}x\to 1$.

即是说这个积分之值"集中"在 $x=0$ 处,这在物理上与第二章§4讲的质量分布是一致的.但这会给人一种印象,即 $\delta(x)$ 的几何图形总是"钟形曲线".然而这是不对的.下一章就会看到,例如还有极重要的费耶(Fejer,匈牙利数学家)核 $K_n(x)=\frac{1}{n}\left[\frac{\sin(n\pi x)}{\sin\pi x}\right]^2$,它在傅里叶级数理论中极为重要.我们将在下一章解释何以会有这种类型的 $\delta$ 序列.到那时就可以进一步体会到广义函数理论的深刻性.

适合以上(i),(ii),(iii)的 $\delta$ 序列时常称为正型序列.

# §7 复 积 分

**1. 复变量函数的积分** 在讨论了实变量函数的微分学以后,我们很自然地进入了复域.这时就发现了:在我们面前展现了一片新天地:即解析函数类.现在我们也要把积分学推广到复域中,令我们吃惊的是:我们又一次遇到了解析函数类.这样我们终于认识到,复变量函数的微积分是一个新的领域,它有自己的问题和方法.因此,这一节里我们关心的不再是函数的可积性问题,从而可以限制于黎曼积分的讨论,而把注意力放在由于进入复域而产生的新问题.我们特别关心的将是如何对待、处理一些拓扑学的问题.拓扑学在现代数学的理论和应用中占据特别重要的位置.本书当然不可能介绍拓扑学,但是在讨论复域中的积分时不利用机会介绍一些拓扑学的思想,是令人遗憾的.

正如在实的情况下我们习惯把一个一元函数写成 $y=f(x)$ 一样,在复情况下,自变量常以 $z=x+\mathrm{i}y$ 表示,函数值以 $w=u+\mathrm{i}v$ 表示,本书中凡见到 $w=f(z)$ 就一定是指复变量的函数.作为复函数,它是一元函数,作为实函数,它是很特殊的二元(实自变量)函数;因为 $z=x+\mathrm{i}y$,故 $\bar{z}=x-\mathrm{i}y$,而 $x=\frac{1}{2}(z+\bar{z})$,$y=\frac{1}{2\mathrm{i}}(z-\bar{z})$,所以一个一般的二元实自变量函数 $f(x,y)$ 将写为

$f(x,y)=f\left[\frac{1}{2}(z+\bar{z}),\frac{1}{2\mathrm{i}}(z-\bar{z})\right]=F(z,\bar{z})$. 这样 $w=f(z)$ 只能是一类特殊的二元实自变量(但可能是复值)函数(即相应的 $F(z,\bar{z})$ 不依赖于 $\bar{z}$)了. 本章将一直使用这样的记号.

实域中通常的定积分 $\int_a^b f(x)\mathrm{d}x$ 之积分区域是 $\mathbf{R}^1$ 的一个区间 $[a,b]$,但是 $\mathbf{R}^1$ 可以看成是复平面 $\mathbf{C}^1$ 中的一条曲线. 因此,要把它推广到复域中就有两种考虑:一是考虑 $\mathbf{C}^1$ 中某一曲线 $L$ 上的积分. 二是考虑 $\mathbf{C}^1$(作为 $\mathbf{R}^2$ 看待)的一个 2 维子区域 $\Omega$ 上的二重积分 $\iint_\Omega f(x,y)\mathrm{d}x\mathrm{d}y$. 但在两种情况下,都需要对 $L$ 或 $\partial\Omega$ 之边界的光滑性加以限制.

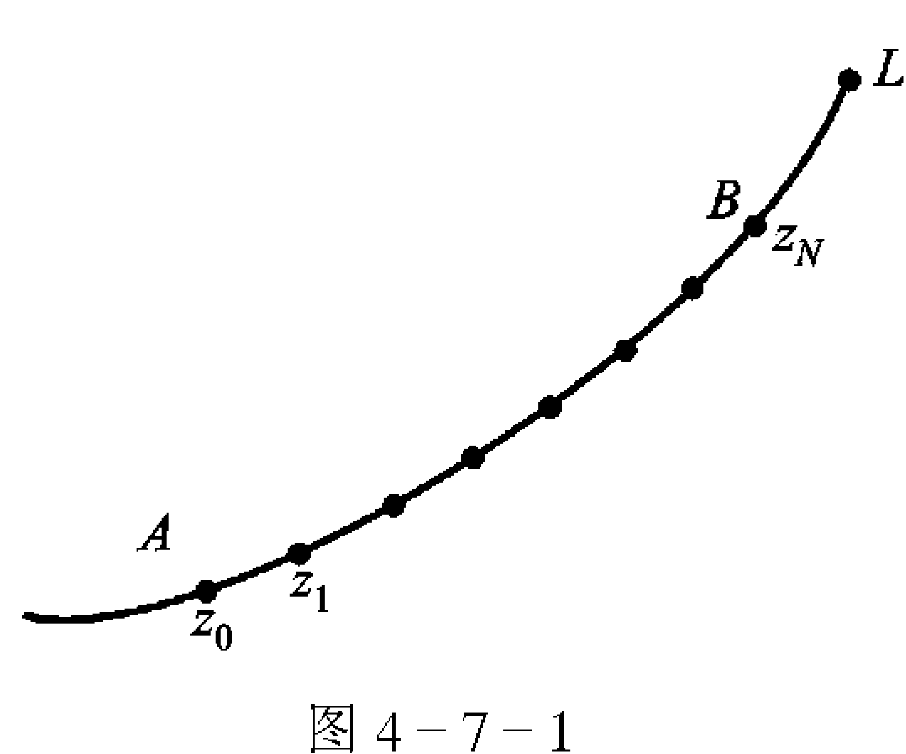



图 4-7-1

先看曲线 $L$,设其参数方程为

$$x=x(t),\ y=y(t). \tag{1}$$

如果要在 $L$ 上的 $AB$ 弧段上积分,仿照黎曼积分的作法,先选一些分点 $A=z_0,\ z_1,\cdots,z_N=B$(图 4-7-1),($z_j$ 的相应参数值为 $t_j$),然后作积分和

$$\sum_{j=0}^{N-1} f(\xi_j)(z_{j+1}-z_j), \tag{2}$$

再求极限. 因为 $z_j=x_j+\mathrm{i}y_j$,

$$\Delta z_j=z_{j+1}-z_j=(x_{j+1}-x_j)+\mathrm{i}(y_{j+1}-y_j),$$

所以

$$|x_{j+1}-x_j|,|y_{j+1}-y_j|\leqslant|\Delta z_j|=\sqrt{|x_{j+1}-x_j|^2+|y_{y+1}-y_j|^2}.$$

一个便利的方法是设 $x=x(t),y=y(t)\in C^1$,于是

$$|\Delta z_j|\leqslant\max_{t_j<t\leqslant t_{j+1}}\sqrt{x'(t)^2+y'(t)^2}(t_{j+1}-t_j) \tag{3}$$

这样求积分和的极限就化为对实变量 $t$ 的分割 $t_0<t_1<\cdots<t_N$ 求积分和的极限,即需令 $\max\limits_j|t_{j+1}-t_j|\to 0$ 并求极限.

当 $x(t),y(t)\in C^1$ 时,曲线弧 $\widehat{AB}$ 可求长:

$$S_{AB}=\int_{t_0}^{t_N}\sqrt{x'(t)^2+y'(t)^2}\,\mathrm{d}t. \tag{4}$$

但是这只是 $\widehat{AB}$ 可求长的一个充分条件. 常用的比较更广泛的可求长条件是,相应于上述分割可以得到

$$\mathrm{V}_P x(t)=\sum_{j=0}^{N-1}|x(t_{j+1})-x(t_j)|,$$

$$\mathrm{V}_P y(t)=\sum_{j=0}^{N-1}|y(t_{j+1})-y(t_j)|. \tag{5}$$

如果这两个量对一切可能的分割 $P$ 有有限的上确界

$$\mathrm{V}x(t)=\sup_P \mathrm{V}_P x(t)<+\infty,\ \mathrm{V}y(t)=\sup_P \mathrm{V}_P y(t)<+\infty, \tag{6}$$

就可以定义$\overset{\frown}{AB}$之长为

$$S_{AB}=\mathrm{V}|z(t)|=\sup_{P}\mathrm{V}_P\sqrt{x(t)^2+y(t)^2}. \tag{7}$$

这时我们就简单地说 $L$ 是可求长曲线.

如果一个函数 $\varphi(t)$，$t_0\leqslant t\leqslant t_{N-1}$ 适合(6)那样类型的条件：

$$\sup_{P}\mathrm{V}_P\varphi(t)=\sup\sum_{j=0}^{N-1}|\varphi(t_{j+1})-\varphi(t_j)|<+\infty, \tag{8}$$

就称之为有界变差函数，所以，可求长曲线就是具有某种参数表示(1)，其中 $x(t)$，$y(t)$均为有界变差的函数的曲线.关于有界变差函数详见§3.

但是在以下凡讲到曲线 $L$ 以及 $\Omega$ 的边界为曲线时，总是设它们可分为有限段，而每一段均为 $C^1$ 曲线弧，有时我们会用到封闭曲线，而且假设它是若尔当曲线.或简单曲线，也就是不自交的封闭曲线；也就是其参数表示(1)中的 $x(t)$，$y(t)$ 不会有 $(x(t_1^*),y(t_1^*))=(x(t^{**}),y(t^{**}))$——除非 $t^*=t_0$，$t^{**}=t_N$——的曲线，这时可以应用著名的若尔当定理：平面上的任一若尔当曲线必将该平面分为两部分(内部和外部)，把某一部分的某点 $P$ 与到另一部分的某点 $Q$ 连结起来的连续曲线必定与该若尔当曲线相交.这虽然是一个著名的应该证明的结果，但我们不再引述其冗长的证明了(图 4－7－2).

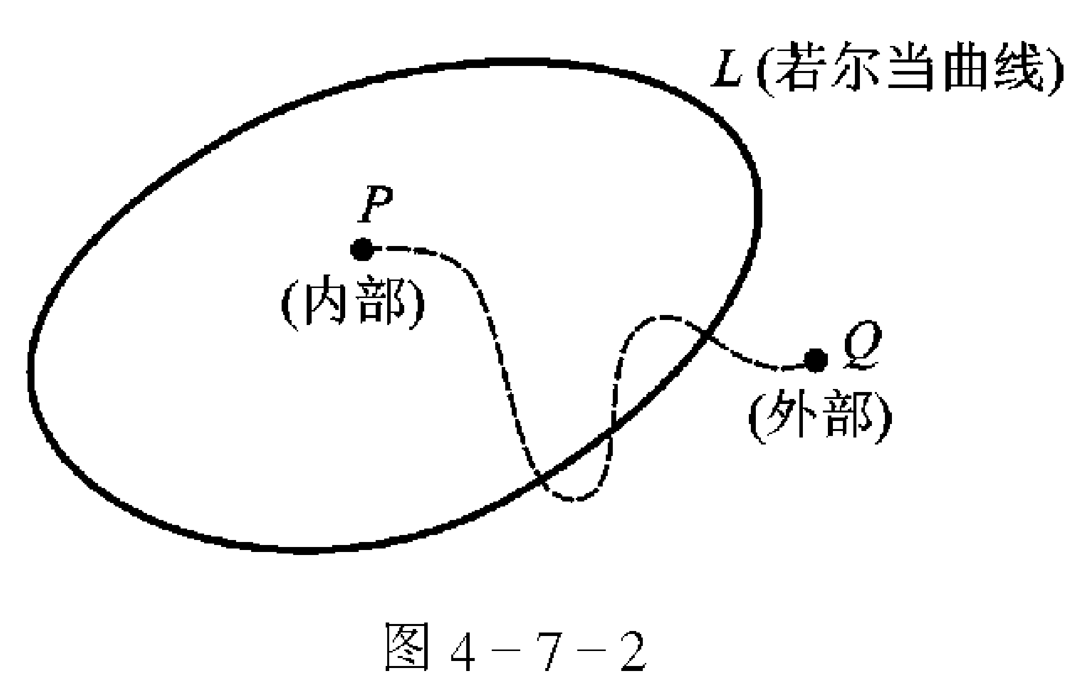


图 4－7－2

复积分的另一个情况是考虑 $f(x,y)$在复平面的某一个区域 $\Omega$ 上的二重积分$\iint\limits_{\Omega}f(x,y)\mathrm{d}x\mathrm{d}y$.由§1 所述，讨论二重积分时需要假设 $\Omega$ 若尔当可测.为此应要求 $\Omega$ 之边界$\partial\Omega$ 具有 0 勒贝格测度.若我们已假定了$\partial\Omega$ 是一个分段(其实是分有限段)$C^1$ 曲线，这个条件是自然会满足的.这是我们约定恒假设$\partial\Omega$ 为分段 $C^1$ 曲线的原因.我们既然要讨论复积分，最好是使用复记号.上面我们已说了 $f(x,y)$可写为 $F(z,\bar{z})$，为简单起见以下我们就直接讨论 $f(z,\bar{z})$在 $\Omega$ 上的积分.$\mathrm{d}x\mathrm{d}y$ 也有复记号：我们把它写成$\frac{\mathrm{i}}{2}\mathrm{d}z\wedge\mathrm{d}\bar{z}$.“$\wedge$”称为外积，关于外积及其与 $\mathrm{d}x\mathrm{d}y$ 的关系，本书最后一章要详细讨论，我们目前只要记住：第一，两个向量 $A$ 和 $B$ 的外积，很像通常微积分教本中讲的向量积，其大小即以这两个向量为边的平行四边形面积，所以 $\mathrm{d}z\wedge\mathrm{d}\bar{z}$“是”一个面积.这自然十分适合二重积分的需要.第二：外积运算适合分配律和反交换律：

$$A\wedge(B+C)=A\wedge B+A\wedge C,$$
$$A\wedge B=-B\wedge A，所以\ A\wedge A=0.$$

我们目前只需把它们当作一种形式规律，而

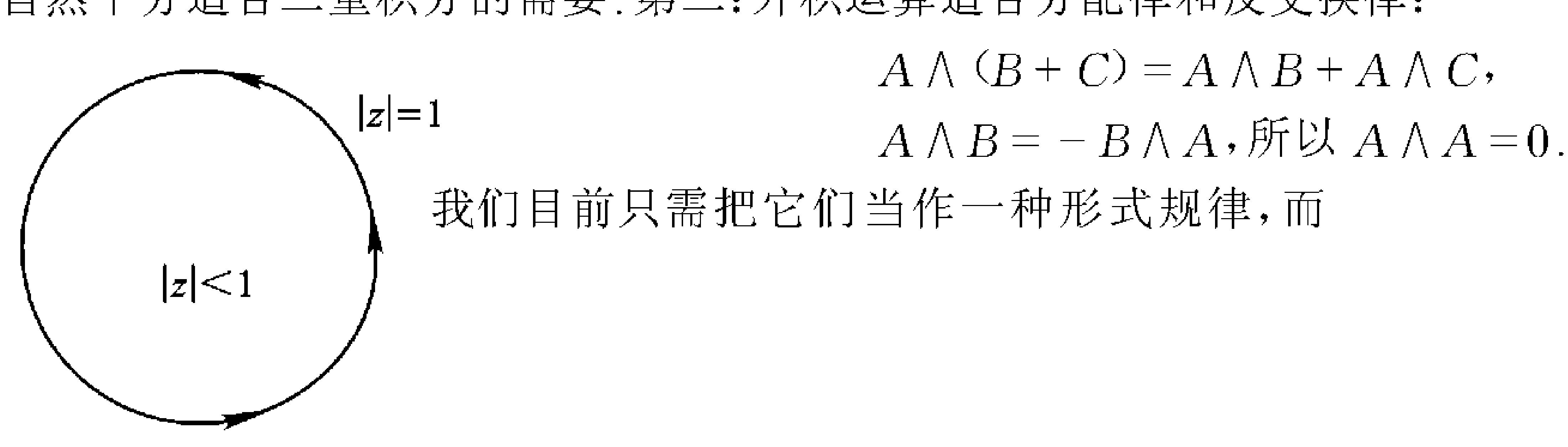


图 4－7－3

它写成$\frac{\mathrm{i}}{2}\mathrm{d}z\wedge\mathrm{d}\bar{z}$.

最后还有一点要提醒,在讨论某一简单封闭曲线 $L$ 上的积分时,时常要在 $L$ 上规定一个"定向"——通常讲的顺、逆时针方向与左、右手坐标系都是讲的"定向"问题——$L$ 上的定向是相对于 $L$ 所围的区域而言的,因为前面介绍的若尔当定理已指出,$L$ 把整个复平面分成(即"围成")两个区域,现在指定其中一个,使当一点沿 $L$ 运动时此区域恒在左方,就说此方向对于所指定的区域是正向.例如单位圆周对于单位圆而言逆时针方向就是正向(图 4-7-3).$L$ 上规定了一个正向后就说 $L$ 是有定向的.但有时也要讨论 $L$ 不是简单曲线的情况,这时 $L$ 上的"正向"就纯属一种约定.

现在我们可以讨论复积分了.我们现在讨论第一种情况,即一维的 $L$ 上的曲线积分,即要求一个连续函数 $f(z)=u+\mathrm{i}v$ 沿一分段 $C^1$ 曲线 $L$ 由 $A$ 点到 $B$ 点的积分.这里 $L$ 是有定向的,而由 $A$ 到 $B$ 即为正向(将定向倒转为由 $B$ 到 $A$ 为正向的 $L$ 则记作 $-L$).我们给出

**定义 1**　$f(z)$沿有向路径 $L$ 由 $A$ 到 $B$ 的积分定义为

$$\int_L f(z)\mathrm{d}z=\int_A^B f(z)\mathrm{d}z=\int_A^B(u+\mathrm{i}v)(\mathrm{d}x+\mathrm{i}\mathrm{d}y)$$
$$=\int_A^B u\mathrm{d}x-v\mathrm{d}y+\mathrm{i}\int_A^B v\mathrm{d}x+u\mathrm{d}y. \tag{10}$$

式右的积分是通常的实曲线积分.

当然我们也可以仿照实积分的方法定义此积分为积分和的极限

$$\int_A^B f(z)\mathrm{d}z=\lim\sum_{j=0}^{N-1}f(\xi_j)(z_{j+1}-z_j).$$

这个定义与上面的定义 1 显然是一致的.

在我们关于 $L$ 的光滑性和 $f(z)$的光滑性的假设下,(10)中的积分显然是存在的,而且就是通常的黎曼积分.所以下面不再讨论可积性问题.

如果 $L$ 有参数式(1),则因其中的 $x(t)$、$y(t)$均为分段 $C^1$ 的,而 $A,B$ 两点分别相应于参数值 $t=a$ 和 $t=b$,若记 $z'(t)=x'(t)+\mathrm{i}y'(t)$,由(10)式立即有:

$$\int_L f(z)\mathrm{d}z=\int_a^b f(z(t))z'(t)\mathrm{d}t. \tag{11}$$

现在我们将参数变为 $t=t(\tau)$,而 $a=t(\alpha)$,$b=t(\beta)$.并且也设 $t=t(\tau)$为分段 $C^1$ 函数,则由上式右方作变量变换又得

$$\int_L f(z)\mathrm{d}z=\int_\alpha^\beta f[z(t(\tau))]z'(t)t'(\tau)\mathrm{d}\tau$$
$$=\int_\alpha^\beta f[z(t(\tau))]z'(\tau)\mathrm{d}\tau.$$

其形式与(11)是一样的.可见积分 $\int_L f(z)\mathrm{d}z$ 之值与参数选取的方法无关.我们特别要作一个变换 $t=-\tau$,则参数 $\tau$ 之值在 $-b\leqslant\tau\leqslant-a$ 中,$\tau$ 由小到大变化就相当于 $z$ 沿反向变化,或者说 $z$ 在 $-L$ 上变化.由上所述有

$$\int_{-L}f(z)\mathrm{d}z=\int_{-b}^{-a}f[z(-\tau)](-z'(-\tau))\mathrm{d}\tau.$$

但上式右方是复值函数的实变量积分，所以可以应用实变量积分的变量变换公式，把 $-\tau$ 改写为 $t$ 而有

$$\int_{-L} f(z)\mathrm{d}z = \int_b^a f[z(t)]z'(t)\mathrm{d}t = -\int_a^b f[z(t)]z'(t)\mathrm{d}t$$

$$= -\int_L f(z)\mathrm{d}z. \tag{12}$$

这是一个很重要的性质，只在 $L$ 上有了定向以后才会有它成立.

积分(10)的许多性质，例如对被积函数的线性性质以及对积分路径的可加性都和实黎曼积分一样.现在我们要讨论关于积分(10)的一个估计. $L$ 的参数式(11)中的参数 $t$ 自然是实的，如果 $L$ 的定向对应于 $t$ 由小变大，则由(11)有

$$\left|\int_L f(z)\mathrm{d}z\right| \leqslant \int_a^b |f(z)||z'(t)|\mathrm{d}t, \qquad a\leqslant b,$$

但 $|z'(t)|^2 = x'(t)^2 + y'(t)^2$，从而当 $\mathrm{d}t>0$ 时，$|z'(t)|\mathrm{d}t = \mathrm{d}s = |\mathrm{d}z|$. 所以我们有

$$\left|\int_L f(z)\mathrm{d}z\right| \leqslant \int_L |f(z)||\mathrm{d}z| = \int_a^b |f(z)|\mathrm{d}s.$$

这里出现了对弧长的积分，这是一种与(10)不同的曲线积分，其定义为

$$\int_L f(z)\mathrm{d}s = \int_L f(z)|\mathrm{d}z| = \int_L f[z(t)]|z'(t)|\mathrm{d}t.$$

而且规定参数 $t$ 由小变大，这种积分与(10)不同之处在于它不区别 $L$ 与 $-L$：

$$\int_{-L} f(z)\mathrm{d}s = \int_{-L} f(z)|\mathrm{d}z| = \int_{a\leqslant t\leqslant b} f[z(t)]|z'(t)|\mathrm{d}t.$$

$$= \int_L f(z)\mathrm{d}s. \tag{13}$$

不论如何取参数，最后成为通常的定积分时总是由小的参数值积到大的参数值.我们称它是在“不定向”的 $L$ 上的曲线积分.许多常用的微积分教材中则把它称为“第一类曲线积分”而把前面讲的称为第二类曲线积分.其区别在于前者适合(13)式而后者适合(12)式.由于通常的一维定积分是在 $x$ 轴上积分的，$x$ 轴自然是已定向的，而且我们都“默认”其正向是 $x$ 由小变大——向右，所以它是“第二类积分”.但通常的二重积分 $\iint_\Omega f(x,y)\mathrm{d}x\mathrm{d}y$ 由于我们也“默认”了 $\mathrm{d}x\mathrm{d}y$ 作为面积单元总是非负的.因而无所谓定向，即无所谓 $-\Omega$ 上的积分.所以是在“不定向”区域上积分，是“第一类积分”.这个区别反映在它们的变量变换公式上，对前者我们有

$$\int_a^b f(x)\mathrm{d}x = \int_\alpha^\beta f(x)x'(t)\mathrm{d}t, \quad a = x(\alpha), b = x(\beta).$$

$x'(t)$就相当于(12)式中的 $z'(t)$.对后者则有

$$\iint_\Omega f(x,y)\mathrm{d}x\mathrm{d}y = \iint_\omega f[x(s,t),y(s,t)]\left|\det\frac{\partial(x,y)}{\partial(s,t)}\right|\mathrm{d}s\mathrm{d}t.$$

$\left|\det\frac{\partial(x,y)}{\partial(s,t)}\right|$则相应于 $|z'(t)|$.第二类积分可以化为第一类积分如下

$$\int_L f(z)\mathrm{d}z = \int_a^b f[z(s)](\mathrm{sgn}\, L)|z'(s)|\mathrm{d}s, \ a\leqslant b.$$

这里出现因子 sgn $L$：当 $L$ 的正向与 $s$ 由小到大的方向一致时就取为 $+1$，否则就取为 $-1$. 当然，这只是形式上的改变；把 $L$ 定向的改变转移到因子 sgn $L$ 上去了.

但是这究竟是一个很重要的区别. 第七章中还要仔细讨论. 本节中我们则只限于讨论第二类的积分(10). 由于它的实部和虚都是形为

$$\int_L P(x,y)\mathrm{d}x+Q(x,y)\mathrm{d}y \tag{14}$$

的积分，我们就从讨论它开始.

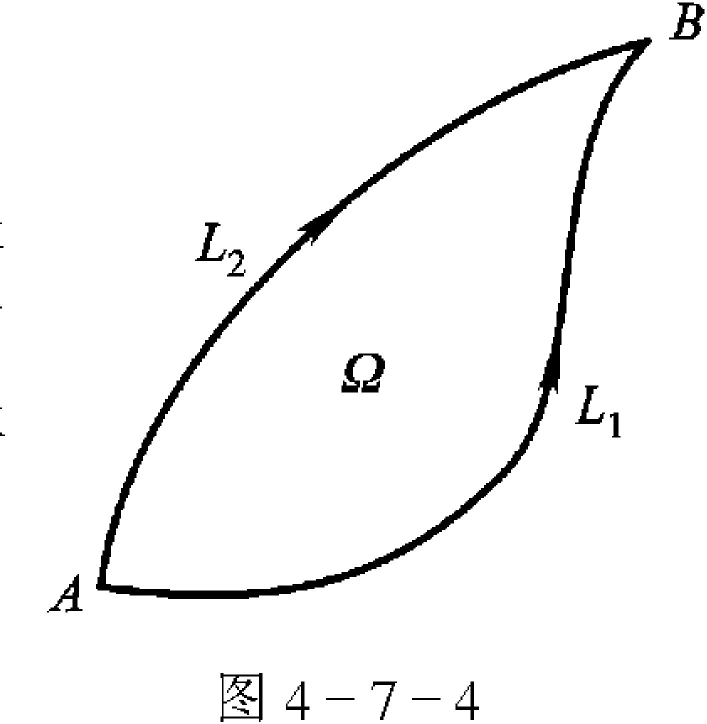



图 4-7-4

**2. 柯西定理** 如果 $L=\overset{\frown}{AB}$的两端一端$A$ 固定，另一端 $B$ 为 $z$ 可以变动，则积分路径 $L$ 的选择，将影响积分(14)之值. 宁可说积分(14)是 $L$ 的泛函，而不能说(14)定义了一个 $z$ 的函数. 为了使此积分定义 $z$ 的函数就需要它与积分的路径无关. 在通常的微积分教材中都介绍了格林公式

$$\int_{\partial\Omega} P\mathrm{d}x+Q\mathrm{d}y=\iint_{\Omega}\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\mathrm{d}x\mathrm{d}y.$$

这里$\partial\Omega$ 对于$\Omega$ 取正向. 如果(14)与 $L$ 无关，作两条 $L_1, L_2$ 连接 $A, B$，如果由 $A$ 出发沿$L_1$ 到 $B$ 再由 $B$ 沿 $-L_2$ 返回到 $A$ 形成一个闭路包围了一个区域$\Omega$，即 $L_1-L_2=\partial\Omega$(图 4-7-4)，而若 $P, Q$ 在$\Omega\cup\partial\Omega=\bar{\Omega}$上相当正规，则积分(14)与 $L$ 无关的充分必要条件是

$$\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}=0. \tag{15}$$

上面我们说的"相当正规"是指 $P$ 和$Q$ 在$\Omega\cup\partial\Omega=\bar{\Omega}$上属于 $C^1$，其目的是为了使格林公式适用，但这显然是过强的条件. 实际上，只要 $P$ 和$Q$ 在 $\bar{\Omega}$ 上连续而在 $\Omega$ 上具有一阶连续导数，$\partial\Omega$ 是简单可求长曲线即可. 但是证明却十分困难，而对理解事物本质助益不大，所以我们宁取"相当正规"的模糊说法以绕过困难.

在条件(15)之下积分(14)确实成了 $z=(x,y)$ (即 $B$ 点)的函数 $\Phi(x,y)$了，当 $\Omega$ 为单连通区域(什么是单连通我们下面再解释)时，$\Phi(x,y)$是单值函数. 第二章中我们明确指出，凡函数均应为单值的，而对常见的多值函数应如何理解也作了仔细的讨论. 下面我们也将对 $\Omega$ 不是单连通，从而 $\Phi(x,y)$具有多值性问题作出解释. 总之，$P\mathrm{d}x+Q\mathrm{d}y$ 将成为$\Phi(x,y)$的全微分，于是

$$P=\frac{\partial\Phi}{\partial x},\quad Q=\frac{\partial\Phi}{\partial y}.$$

这样一来就发现条件(15)即$\dfrac{\partial^2\Phi}{\partial x\partial y}=\dfrac{\partial^2\Phi}{\partial y\partial x}$. 这个似乎不引人注目的事在物理上却有很重要的含意. 如果设有静电场 $\boldsymbol{E}=(E_x,E_y)$，而 $P, Q$ 是$\boldsymbol{E}$ 的两个分量，则积分(14)表示将单位电荷由 $A$ 移到$B$ 点电场所作的功. 这个功的大小(如果固定 $A$ 点)只决定于 $B$ 之位置而与由$A$ 到$B$ 的路径选取方法无关. 物理学中有许多场都具有这个性质. 例如静电场与引力场，这种场称为有势场，函数 $\Phi(x,y)$称为势函数(不过在物理学中我们却称 $U(x,y)=-\Phi(x,y)$为电位，它与电场强度的关系是 $\boldsymbol{E}=-\mathbf{grad}\ U$. 对于引力场，$U$ 称为势能. 势能也叫位能，即由位置引起的"蓄势待发"的能). 势函数的存在是一个在数学上和物理上都十分重要的事，本书中将多次在第七章中遇

到这个问题.

现在我们就回到复积分的情况.首先我们引入格林公式的复表述方法.现在我们引入两个形式符号.

$$\frac{\partial}{\partial \bar{z}}=\frac{1}{2}\left(\frac{\partial}{\partial x}+\mathrm{i}\frac{\partial}{\partial y}\right),\ \frac{\partial}{\partial z}=\frac{1}{2}\left(\frac{\partial}{\partial x}-\mathrm{i}\frac{\partial}{\partial y}\right),$$

称为庞贝(Pompeiu)记号(Pompeiu 是罗马尼亚数学家)而 $f(z)=u+\mathrm{i}v$ 为解析的 C-R 条件即

$$\frac{\partial f}{\partial \bar{z}}=0.$$

现在我们有

**定理 1** 若 $f(z)\in C^1(\Omega)$ 而 $\partial\Omega$ 是分段 $C^1$ 曲线,则

$$\int_{\partial\Omega} f(z)\mathrm{d}z=\iint_{\Omega}\mathrm{d}f\wedge\mathrm{d}z. \tag{16}$$

这个定理我们不加证明了,因为它就是对(10)中的两个实曲线积分 $\int_{\partial\Omega} u\mathrm{d}x-v\mathrm{d}y$ 与 $\int_{\partial\Omega} v\mathrm{d}x+u\mathrm{d}y$ 应用实的格林定理的结果.

但若注意到

$$\begin{aligned}\mathrm{d}f&=\frac{\partial f}{\partial x}\mathrm{d}x+\frac{\partial f}{\partial y}\mathrm{d}y=\frac{1}{2}\left[\frac{\partial f}{\partial x}(\mathrm{d}z+\mathrm{d}\bar{z})+\frac{1}{\mathrm{i}}\frac{\partial f}{\partial y}(\mathrm{d}z-\mathrm{d}\bar{z})\right]\\&=\frac{\partial f}{\partial z}\mathrm{d}z+\frac{\partial f}{\partial \bar{z}}\mathrm{d}\bar{z}.\end{aligned}$$

再利用"外积"$\wedge$的反对称性:$\mathrm{d}z\wedge\mathrm{d}z=0$,就可以把(16)化为

$$\int_{\partial\Omega} f(z)\mathrm{d}z=\iint_{\Omega}\frac{\partial f}{\partial \bar{z}}\mathrm{d}\bar{z}\wedge\mathrm{d}z. \tag{17}$$

请读者决不要对此等闲视之.表面上看这里只有一些简单的形式计算,而实际上却揭示了积分深刻地体现了一种对偶性.上一节我们说积分可以看作作用在被积函数上的泛函,因而得出了广义导数乃至广义函数.这是在被积函数空间的对偶空间(即连续线性泛函之空间)上做文章.现在更进一步看到被积函数与积分区域之间也有一种对偶关系:对 $f$ 的求导运算$\frac{\partial}{\partial \bar{z}}$(虽然现在只是一个形式记号)与对 $\Omega$ 的一个拓扑操作——取边缘——$\partial$相对偶.(17)式只是深刻的斯托克斯定理最简单的情况.斯托克斯定理是第七章的重要内容.

现在回到 $f(z)$ 是 $z$ 在 $\Omega\subset C$ 中的解析函数的情况.先设 $\Omega$ 有一个子域 $\Omega_1$,其边缘 $\partial\Omega_1$ 是一个若尔当曲线,其内域即 $\Omega_1$,通常我们设 $\Omega_1$ 为单连通的,单连通的定义下面再给.这时利用格林公式(17)立即可得.

**定理 2(柯西定理,1875)** 若 $f(z)$ 在 $\Omega$ 中解析,$\bar{\Omega}_1\subset\Omega$,$\partial\Omega_1$ 是分段 $C^1$ 的若尔当曲线,而以 $\Omega_1$ 为其内域,则

$$\int_{\partial\Omega_1} f(z)\mathrm{d}z=0.$$

**证明非常简单**：直接利用(17)式即可．问题在于定理的陈述．我们没有说 $\int_{\partial\Omega} f(z)\mathrm{d}z=0$，因为 $f(z)$ 在 $\partial\Omega$ 上是否有定义尚且不知，更说不上是否可积了．如果我们将条件改成"$f(z)$ 在 $\bar{\Omega}$ 上解析"，则由函数解析性的定义可知，它必在 $\bar{\Omega}$ 的某一个邻域中解析．把这个邻域记为 $\Omega$ 而把原来的 $\Omega$ 改写成 $\Omega_1$ 即得．我们对此定理的陈述还大有改进的余地．例如设 $\partial\Omega$ 是分段 $C^1$ 曲线，$f(z)$ 在 $\Omega$ 中解析，而在 $\bar{\Omega}$ 上连续又如何？结果是：定理 2 此时仍是对的．实际上，现在通用的许多关于复分析的书籍中，这个定理的条件都放宽了，因为人们后来把 $f(z)$ 在 $z_0$ 点解析的定义改成了：$f(z)$ 在 $z_0$ 的某一邻域中可求导，而不要求 $f'(z)$ 的连续性．这样一来就不能再用格林公式了．古尔萨(E. Goursat)在解析性的这个定义下不用格林公式证明了如我们所陈述的定理，所以在有的书上称定理 2 为柯西－古尔萨定理．最后又经许多人努力才把条件放松到"$\Omega$ 中解析，$\bar{\Omega}$ 上连续，且 $\partial\Omega$ 是可求长曲线"．证明复杂多了，但对问题的理解好处有限，所以我们采用现在的讲法．

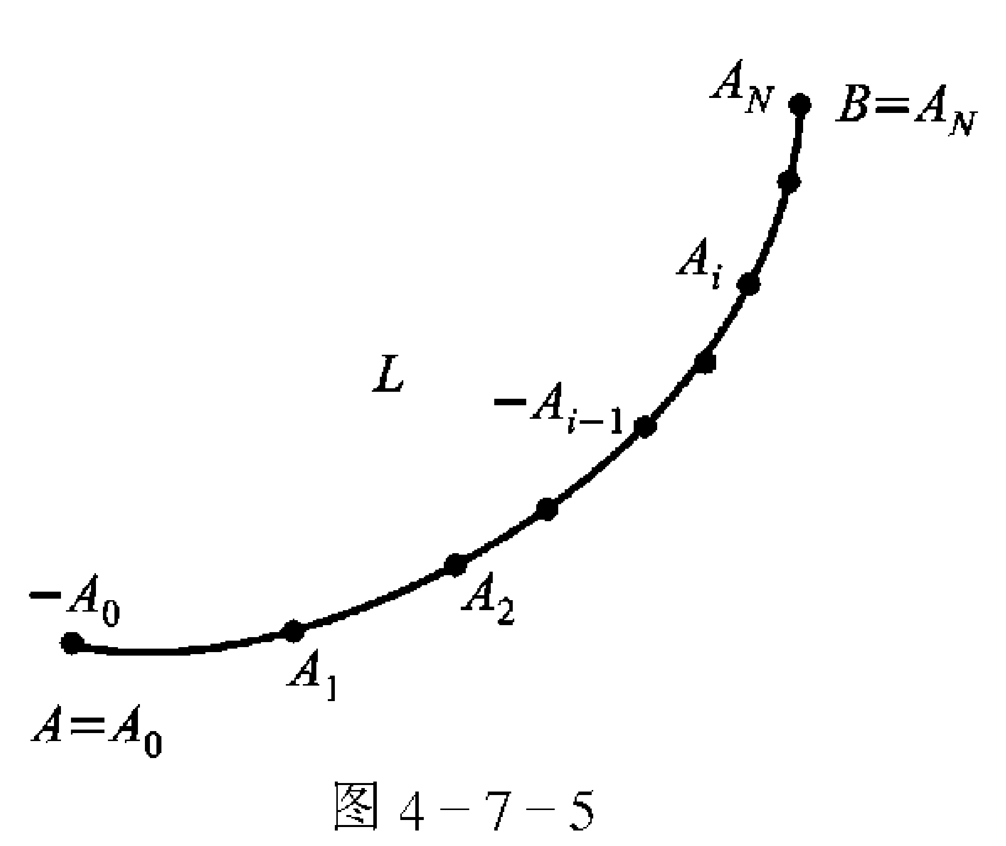


图 4－7－5

真正的问题有二：一是积分路径不一定是若尔当曲线，而可以自交．也不一定是某区域的边缘．对这个问题的处理的方法如下：曲线由曲线段构成，而曲线段是闭间 $I=[0,1]\subset\mathbf{R}^1$ 在某映射 $\varphi: I\to\mathbf{C}$ 下的像(这里 $\mathbf{C}$ 指复数域，$I$ 中的变元就是曲线段的参数)，$I$ 的两个端点(不妨设为 $t=0$，$t=1$)的像构成曲线段的边缘，记作 $\partial(\widehat{AB})$ 不过起点加上负号，终点带上正号，以示该曲线段的定向：$\partial(\widehat{AB})=B-A$．每个曲线段称为一个 1 维单形(simplex)，而一条曲线则由有限多个 1 维单形"首尾相连"而成，我们可以形式地写为

$$
\begin{aligned}
L &= \widehat{AA_1}+\widehat{A_1A_2}+\cdots+\widehat{A_{N-1}B},\ A=A_0,\ B=A_N,\\
\partial L &= \partial(\widehat{AA_1})+\partial(\widehat{A_1A_2})+\cdots+\partial(\widehat{A_{N-1}B})\\
&= (B-A_{N-1})+(A_{N-1}-A_{N-2})+\cdots+(A_2-A_1)+(A_1-A)\\
&= B-A.
\end{aligned}
$$

其中例如 $A_1$ 出现两次，一次取正号表示它是 $\widehat{A_0A_1}$ 之尾，另一次取负号表示它是 $\widehat{A_1A_2}$ 之首，这就是"首尾相连"的意思——因首尾相连，故正负相消．于是，一条曲线称为一个 1 维链(chain)；链是单形的代数和．如果一条曲线是封闭的，即上面的 $A$ 与 $B$ 相同而又首尾相连，从而 $\partial L=0$，就称它是一个闭链或循环(cycle)．所以积分区域 $L$ 就是一个闭链．现在我们把 $L$ 这个闭链分解如下：如果 $L$ 有自交点，我们可以把 $L$ 分成几段，使每一段都从一个自交点开始并再回到此点，但此段之中不再有其它自交点．于是每一段都是一个闭链：

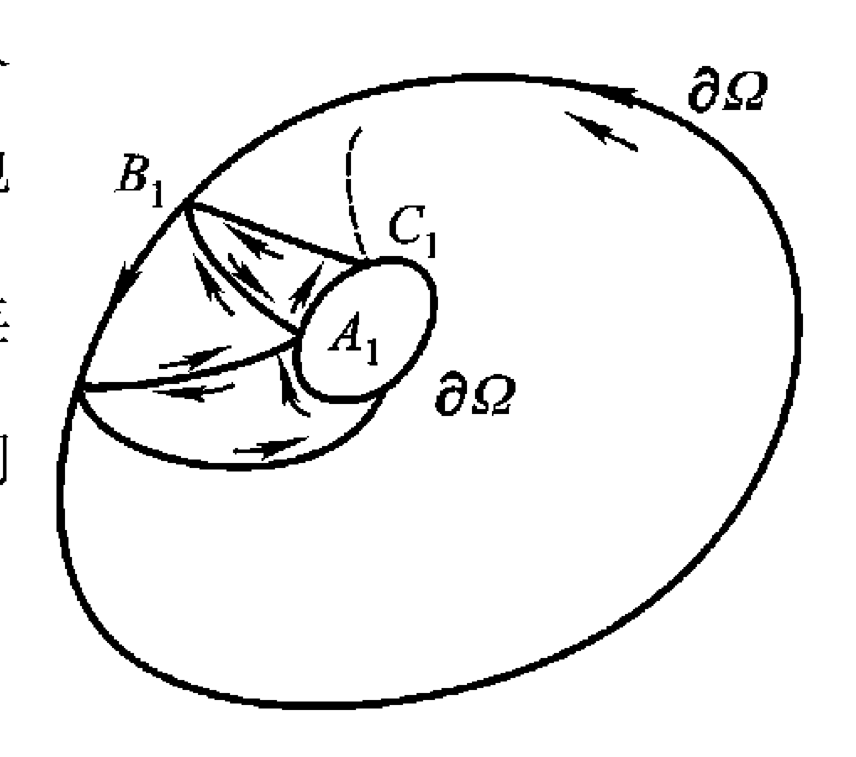
$$L=L_1+L_2+\cdots+L_k.$$

然后就有

$$\int_L f(z)\mathrm{d}z=\sum_{j=1}^{k}\int_{L_j} f(z)\mathrm{d}z,$$

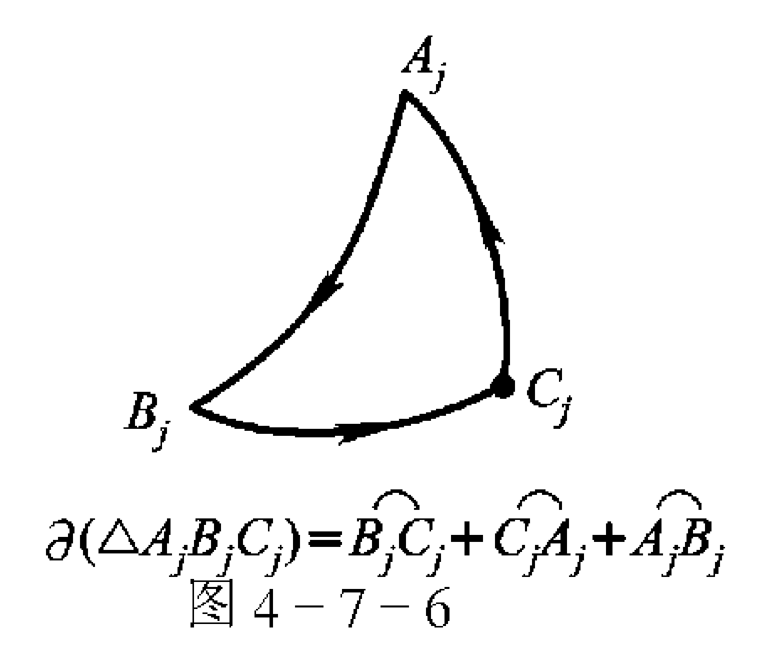
而每一个 $L_j$ 都是一条封闭的若尔当曲线.

第二个问题是看 $L$ 是否包围一个区域 $\Omega$. 为此我们又把 $\Omega$ 分成 $k$ 个小的曲边三角形 $\triangle A_jB_jC_j\ (j=1,2,\cdots,k)$，所以我们形式地把它记作 $\Omega=\sum\limits_{j=1}^{k}\triangle A_jB_jC_j$ 并称之为一个 2 维链(图 4-7-6). 每个这样的三角形称为一个 2 维单形. 对每个三角形 $\triangle A_jB_jC_j$，我们定义其边缘是




$$\partial(\triangle A_jB_jC_j)=\overset{\frown}{B_jC_j}+\overset{\frown}{C_jA_j}+\overset{\frown}{A_jB_j}.$$

边缘的定向的要求是：其定向必须与 $\partial\Omega$ 原有的定向相容. 因为把 $\Omega$ 划分成三角形是利用 $\partial\Omega$ 再加上一些曲线段形成的. 在最简单的情况下，后加上的曲线段一定位于两个三角形共同的边缘上. 这两个三角形一在其左. 一在其右. 所以这个加上的曲线一定要走两次且方向相反：第一次保持某一个三角形在其左方从而定义了这一次走的是正向. 第二次一定要走反向以保持另一个三角形在左方. 这样一来，如果原来的 $\Omega$ 被划分成了 $\Omega_1+\Omega_2+\cdots+\Omega_k$，$\Omega_j=\triangle A_jB_jC_j$ 则




$\partial(\triangle A_jB_jC_j)=\overset{\frown}{B_jC_j}+\overset{\frown}{C_jA_j}+\overset{\frown}{A_jB_j}$

图 4-7-6

$$\partial\Omega=\sum_{j=1}^{k}\partial(\triangle A_jB_jC_j).$$

一个区域的边缘一定是一个闭链；因为边缘再没有边缘了：例如一个圆盘的边缘是圆周，但圆周再没有边缘了，或者说其边缘为 0. 因此

$$\partial^2\Omega=\partial L=0.$$

总之，我们有

$$\partial^2=0.$$

这是一个很深刻的关系式，将来我们会看到，它与微分学的一个基本公式 $\dfrac{\partial^2 f}{\partial x\partial y}=\dfrac{\partial^2 f}{\partial y\partial x}$ 是互相对偶的.

至此为止，我们都是在作一些形式运算. 读者会问，边缘为零何不直接说是边缘为一空集合？为什么要说“单形之和”，而不说是一些单形的并集？我们正是有意不用集合的语言，而把几何问题用组合的方法化为一些形式运算. 这样就可以用代数学的方法去研究它. 这个观点是 19 世纪末由法国大数学家庞加莱首创的，它是代数拓扑学或称组合拓扑学的开始，是现代数学一个重要的支柱，而在积分学中开始介绍它是最自然的办法.

在这样处理了这两个问题之后，我们将得到

**定理 3**(**柯西定理的一般形式**)　设 $f(z)$ 在 $\Omega\subset\mathbf{C}$ 中解析，$\bar{\Omega}_1\subset\Omega$ 有分段 $C^1$ 边缘，则

$$\int_{\partial\Omega_1}f(z)\,\mathrm{d}z=0.$$

**证**　将 $\Omega_1$ 写成一个 2 维链如前所述. 为此只需在 $\Omega_1$ 中另加进有限多条曲线段，于是 $\Omega_1=\sum\limits_{j=1}^{k}\omega_j$ 而每个 $\omega_j$ 都是一个 2 维单形，即一个曲边三角形. 对于每一个 $\omega_j$ 都可以应用单连通情况下的柯西定理，即定理 2，于是

$$\int_{\partial\omega_j} f(z)\,dz=0.$$

我们来分析每一个$\partial\omega_j$. $\omega_j$ 作为一个曲边三角形，其边缘是三个 1 维单形，而且各赋有正向——即使得 $\omega_j$ 在其左侧的定向——这些单形可以分成两类：一类是$\partial\Omega_1$ 的一部分，另一类是外加的曲线段. 关于第一类，容易看到整个$\partial\Omega_1$ 都被分成若干段，而每一段都在某个曲边三角形的边缘上，所以第一类 1 维单形加成一个 1 维链$\partial\Omega_1$，它当然是一个闭链. 再看第二类，它们来自外加的曲线，所以每一个都是两个相邻曲边三角形之公共边. 这两个三角形分别位于其左右侧，所以任取一个 1 维单形 $\sigma$，它赋有正向后使某一曲边三角形$\triangle_1$ 在其左侧. 这样，与它相邻的曲边三角形$\triangle_2$ 必使 $-\sigma$ 在其边缘之中. 这里一定要写 $-\sigma$，因为这样才能使$\triangle_2$ 在其左侧. 所以第二类 1 维单形必成对出现，一为 $\sigma$，另一为 $-\sigma$，这样就有

$$\begin{aligned}\sum\partial\omega_j &= \sum_{\text{第一类}}(\text{第一类单形})+\sum_{\sigma\in\text{第二类}}(\sigma-\sigma)\\ &=\partial\Omega_1.\end{aligned}$$

因此我们有

$$\int_{\partial\Omega_1} f(z)\,dz=\int_{\sum\partial\omega_j} f(z)\,dz=\sum_j\int_{\partial\omega_j} f(z)\,dz=0\ .$$

定理证毕.

定理 3 包含了多连通区域的柯西定理. 那么什么叫做单连通，什么叫多连通呢？如果我们不想多用拓扑学的语言而只来看 $\Omega$ 是有界连通区域的情况，所谓 $\Omega$ 为单连通区域就是说$\partial\Omega$ 把复平面分成两块，一块包含无穷远点，不妨称为外域，一块则不包含无穷远点，称为内域. 如果$\partial\Omega$ 把平面分成 $n+1$ 块，则说 $\Omega$ 是 $n$ 连通的. 这 $n+1$ 块中只有一块含无穷远点，称为外域，其它 $n$ 块就可能是 $n$ 个"洞". 所以，多连通区域就是有洞的区域. 每一个"洞"都有自己的边缘 $\Gamma_1,\Gamma_2,\cdots,\Gamma_n$. 外域也有边缘记为 $\Gamma_0$. 它们都按以前的规定有各自的正向. 例如 $\Gamma_0$ 的正向即是把无穷远点置于其右侧的方向，即逆时针向；$\Gamma_1,\cdots,\Gamma_n$ 的正向则是顺时针方向(图 4-7-7). 我们再一次指出，这种表达虽然直观却是不严格的. 但是下面的概念却是严格的，而且具有基本的重要性.

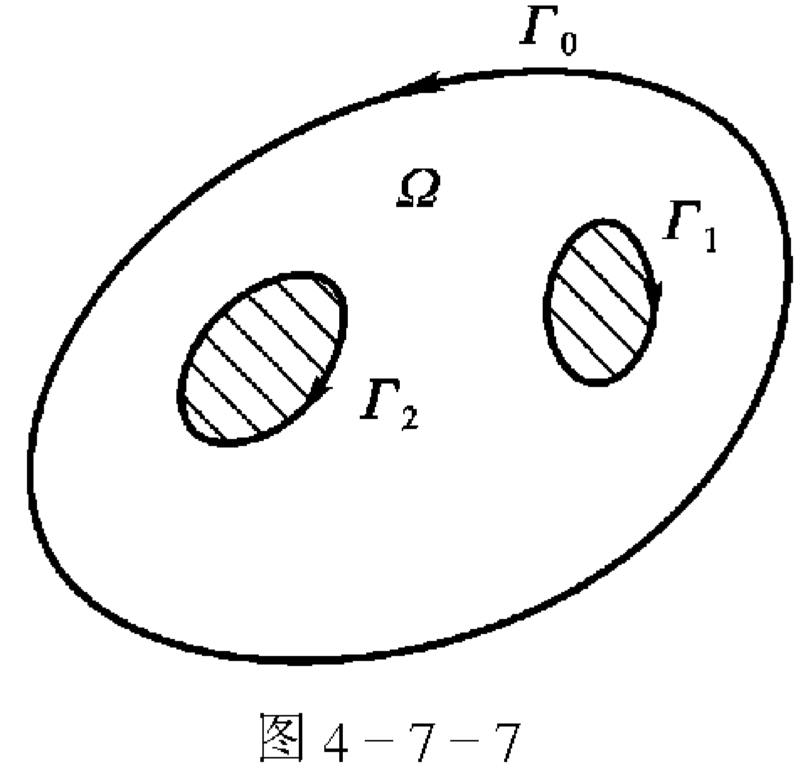


图 4-7-7

**定义 2** 若 1 维闭链 $\gamma$ 是一个 2 维链$\Omega$ 之边缘，即 $\gamma=\partial\Omega$，就说$\gamma$ 同调于 0，记作 $\gamma\sim 0$. 若两个 1 维闭链 $\gamma_1$ 与 $\gamma_2$ 之差是 $\Omega$ 的边缘：$\gamma_1-\gamma_2=\partial\Omega$，就说 $\gamma_1$ 与 $\gamma_2$ 同调，记作 $\gamma_1\sim\gamma_2$.

"1 维闭链"一语中的"闭"字，似乎是多余的：既然 $\gamma$ 已是边缘，自然有$\partial\gamma=0$(前面已经说了)，再加闭字岂非多余？确实如此，但是在 $\gamma_1\sim\gamma_2$ 中，我们只知 $\gamma_1-\gamma_2$ 是闭链而对 $\gamma_1,\gamma_2$ 分别是否闭链并不清楚，但是同调关系是用来处理闭链的等价分类的，所以若 $\gamma_1,\gamma_2$ 本身并非闭链，说 $\gamma_1\sim\gamma_2$ 就只是一种方便说法. 所以在定义中必须指出 $\gamma_1,\gamma_2$ 是闭链.

现在回来看柯西定理. 利用上图，我们再引入一些闭链 $\gamma_0,\gamma_1,\cdots,\gamma_k$. 这一次 $\gamma_0=\Gamma_0$，即定向也不变，但 $\gamma_j=-\Gamma_j$，$j=1,2,\cdots,k$，即内部的那些闭链也改成以逆时针方向为正向. 令 $\tilde{\gamma}=\gamma_1+\cdots+\gamma_k$，则有$\partial\Omega=\gamma_0-\gamma_1-\cdots-\gamma_k=\gamma_0-(\gamma_1+\cdots+\gamma_k)=\gamma_0-\tilde{\gamma}$. 所以 $\gamma_0$ 同调于 $\tilde{\gamma}$，而定理

3 即成为

$$\int_{\partial\Omega} f(z)\mathrm{d}z = \int_{\gamma_0} f(z)\mathrm{d}z - \int_{\tilde{\gamma}} f(z)\mathrm{d}z = 0,$$

亦即

$$\int_{\gamma_0} f(z)\mathrm{d}z = \int_{\tilde{\gamma}} f(z)\mathrm{d}z.$$

一般而言，我们有

**定理 4（柯西定理的同调论表述）** 设 $f(z)$ 在 $\Omega$ 中解析，$\gamma_1$，$\gamma_2$ 为 $\Omega$ 中的两个闭链，若 $\gamma_1 \sim \gamma_2$，则

$$\int_{\gamma_1} f(z)\mathrm{d}z = \int_{\gamma_2} f(z)\mathrm{d}z.$$

注意，同调关系是一个等价关系，互相等价的闭链形成一个同调类. 所以定理 4 实际上是说. 对于固定的 $f(z)\mathrm{d}z$，积分值决定于积分路经的同调类. 下面我们还要指出，被积表达式 $f(z)\mathrm{d}z$（我们不说被函数 $f(z)$ 而说被积表达式，其中有深意焉，详见本书第七章）也有某种等价关系，称为上同调关系. 积分值其实依赖于 $f(z)\mathrm{d}z$ 的上同调类. 这叫做斯托克斯定理，详见第七章. 所以定积分实际上表现了同调类与上同调类的对偶关系.

有了柯西定理就可以回答复积分的原函数问题. 一个复变量函数 $f(z)$（暂时不问它是否解析函数）的原函数 $F(z)$ 就是适合 $F'(z) = f(z)$ 的函数. 对于实变量函数 $f(x)$，如果它连续，则一定有原函数，它的具有可变上限的黎曼积分 $\int_a^x f(t)\mathrm{d}t$ 就是一个原函数. 但对复变量 $z$ 情况就不同了. 这时 $\int_a^z f(\zeta)\mathrm{d}\zeta$（$\zeta$ 也是复的）就不只依赖于 $z$，同时还依赖于由 $a$ 到 $z$ 的路径 $L$，所以它不是 $z$ 的函数. 为了讨论 $f(z)$ 的原函数应设此积分与路径无关. 由此我们容易想到，若 $f(z)$ 是 $\Omega$ 中的解析函数，则积分 $\int_a^z f(\zeta)\mathrm{d}z$ 已经与路径无关（当然这时应规定路径只在 $\Omega$ 中变动），它是否为 $f(z)$ 的原函数？讲到这里应该提到，在复域中研究解析函数有两种观点：局部的和整体的. 局部观点如下：首先在一点附近定义 $f(z)$ 的解析性（利用 $f'(z)$ 连续以及 C－R 方程），然后问 $f(z)$ 在 $\Omega$ 这"一点附近"以外是否解析？这时就需要作解析延拓. 如果作了"一切可能的"解析延拓，则得到一个整体的解析函数. 整体地研究它就发现它可能是多值的，为此就要引入黎曼曲面的概念. 具体的解析延拓如何实现？第二章中讲了一些初等的方法如处理 $\sqrt{z}$ 那样，一般地则可以用幂级数——泰勒级数. 现在我们指出，在证明了 $\int_a^z f(\zeta)\mathrm{d}\zeta$ 是 $f(z)$ 的原函数以后，这个积分实际上正是把原函数沿积分路经作解析延拓的具体方法之一. 因此我们先从 $\int_a^z f(\zeta)\mathrm{d}\zeta$ 已经成为 $z$ 的函数 $F(z)$ 开始. 但首先设 $\Omega$ 是一个单连通区域，甚至只是 $|z-a| < r$，它是 $a$ 点的单连通邻域，并在 $\Omega$ 中证明我们的结论. 即我们采取局部的观点. 至于如何把这些结果解析延拓到整体去，最后会得到一个什么样的多值函数，则是另外一回事了. 由此至定理 8 都这样处理，所以下面不再说这样的话了. 今证 $F(z)$ 是 $f(z)$ 的原函数. 为此我们来考察

$\frac{1}{\Delta z}[F(z+\Delta z)-F(z)]$.我们先取一个路径 $L$ 联结 $a$ 与 $z$ 这样得出 $F(z)=\int_L f(z)\mathrm{d}z$.在计算 $F(z+\Delta z)$时,因为积分与路径无关,我们取一个特殊路径 $L_1=L+[z,z+\Delta z]$,这里$[z,z+\Delta z]$表示连接这两点的直线段(当 $|\Delta z|$ 充分小时,这个直线段总在 $\Omega$ 中).所以 $\frac{1}{\Delta z}[F(z+\Delta z)-F(z)]=\frac{1}{\Delta z}\int_{[z,z+\Delta z]}f(\zeta)\mathrm{d}\zeta=f(z)+\frac{1}{\Delta z}\int_{[z,z+\Delta z]}[f(\zeta)-f(z)]\mathrm{d}\zeta$.在计算第一项时我们利用了 $\int_{[z,z+\Delta z]}\mathrm{d}z=\Delta z$,这应该由积分的定义算出,而不能用牛顿-莱布尼茨公式,因为在复域中这个公式是否成立还不得而知.至于第二项,利用 $f(z)$之连续性有

$$\left|\frac{1}{\Delta z}\int_{[z,z+\Delta z]}[f(\zeta)-f(z)]\mathrm{d}z\right|\leqslant\frac{1}{|\Delta z|}\max|f(\zeta)-f(z)|\int_{[z,z+\Delta z]}|\mathrm{d}z|$$
$$=\max_{[z,z+\Delta z]}|f(\zeta)-f(z)|\to 0.$$

由此得到下面的

**定理 5** 单连通区域 $\Omega$ 中的解析函数 $f(z)$ 必有原函数 $F(z)=\int_a^z f(\zeta)\mathrm{d}\zeta$,其中 $a\in\Omega$,而积分路径含在 $\Omega$ 中.

这里要求 $\Omega$ 是连通的是因为这个积分只能在含 $a$ 的连通分支中定义 $F(z)$.至于单连通问题下面再说.

**定理 6** 单连通区域 $\Omega$ 中的解析函数 $f(z)$,若以 $F(z)$ 为其一个原函数,则 $F(z)+C$ 就都是 $f(z)$ 的原函数,而且给出了其全部原函数.

**证** 与实变函数情况一样,我们只需证明若 $F'(z)=0$,则 $F(z)=C$ 即可.但是这里的证法与实的情况不同,因为复的情况下没有拉格朗日公式.这是实的与复的情况的重要区别.现令 $F(z)=u(x,y)+\mathrm{i}v(x,y)$,由于 $F'(z)$之值应与 $z+h$ 趋向 $z$ 的情况无关.令 $h$ 为实数,即得 $F'(z)=u_x+\mathrm{i}v_x$,由 $F'(z)=0$ 即得 $u_x=v_x=0$ 于 $\Omega$ 中.由 C-R 条件又有 $u_y\equiv v_y\equiv 0$,所以 $u$,$v$ 都是常数,定理证毕.

由此立即可得复情况下的牛顿—莱布尼茨公式.

**定理 7** 假设同上,我们有

$$\int_a^b f(z)\mathrm{d}z=F(b)-F(a).$$

柯西定理的逆定理仍成立,实际上我们有

**定理 8(摩列拉(Morera)定理)** 若 $f(z)$ 在 $\Omega$ 中连续,而 $\Omega$ 之每一点 $a$ 都有一个单连通邻域,使在其中 $\int_a^z f(\zeta)\mathrm{d}\zeta$ 均与积分路径无关(或者说在每一点 $a$ 均有一个单连通邻域,在其中沿闭曲线 $L$ 均有 $\int_L f(\zeta)\mathrm{d}\zeta=0$),则 $f(z)$ 必在 $\Omega$ 之每一点 $a$ 的附近为解析的.

**证** 由上所述 $\int_a^z f(\zeta)\mathrm{d}\zeta$ 是 $a$ 附近

涉及原函数的定理 5～8 中均假设了 $\Omega$ 是单连通区域，这是有深刻的原因的. 如果 $\Omega$ 是多连通的(即有洞的)区域，原函数将是多值的. 这在物理学上是有根据的. 因为原函数的求法与一个场的势的求法是很相近的：若设 $f(z)=u(x,y)+\mathrm{i}v(x,y)$ 是解析函数，$F(z)=\Phi+\mathrm{i}\Psi$ 是它的原函数：$F'(z)=f(z)$. 如果我们让 $z$ 是沿 $x$ 轴变化的，则 $F'(z)=\Phi_x+\mathrm{i}\Psi_x$. 于是 $\Phi_x=u$，$\Psi_x=v$. 但由 C－R 条件，$\Psi_x=-\Phi_y$，于是有

$$\Phi_x=u,\ \Phi_y=-v.$$

即是说，$\Phi(x,y)$ 是向量 $(u,-v)$ (相应于函数 $\bar{f}(z)=u-\mathrm{i}v$) 的势函数. 势函数成为多值函数，这在物理上有重要的意义. 为什么在多连通区域中原函数会出现多值性呢？试利用积分式 $F(z)=\int_0^z f(\zeta)\mathrm{d}\zeta$ 把 $F(z)$ 最终解析延拓到区域 $\Omega$，直到不能再继续延拓，这时 $\partial\Omega$ 称为 $F(z)$ 的自然边界(natural boundary)，而 $\Omega$ 中有若干个洞 $\Omega_1,\cdots,\Omega_k$. 我们还设在 $\Omega$ 中可以作一个若尔当曲线 $\Gamma_0$ 包围所有的“洞”，又对每一个洞 $\Omega_j$ 可以作一个若尔当曲线 $\Gamma_j$ 包围它. 如图 4－7－8. 于是对 $a$ 点和 $\Omega$ 中一点 $z$，我们先作一条路径 $\gamma_1$ 由 $a$ 到 $z$. 如果再取另一条路径 $\gamma_2$ 由 $a$ 到 $z$，则 $\gamma_2-\gamma_1$，成为一个闭链. 这个闭链的构造可以十分复杂. 例如图 4－7－8 中的 $\gamma_2-\gamma_1$ 实际上绕过 $\Omega_1$ 和 $\Omega_2$ 各两次. 总之

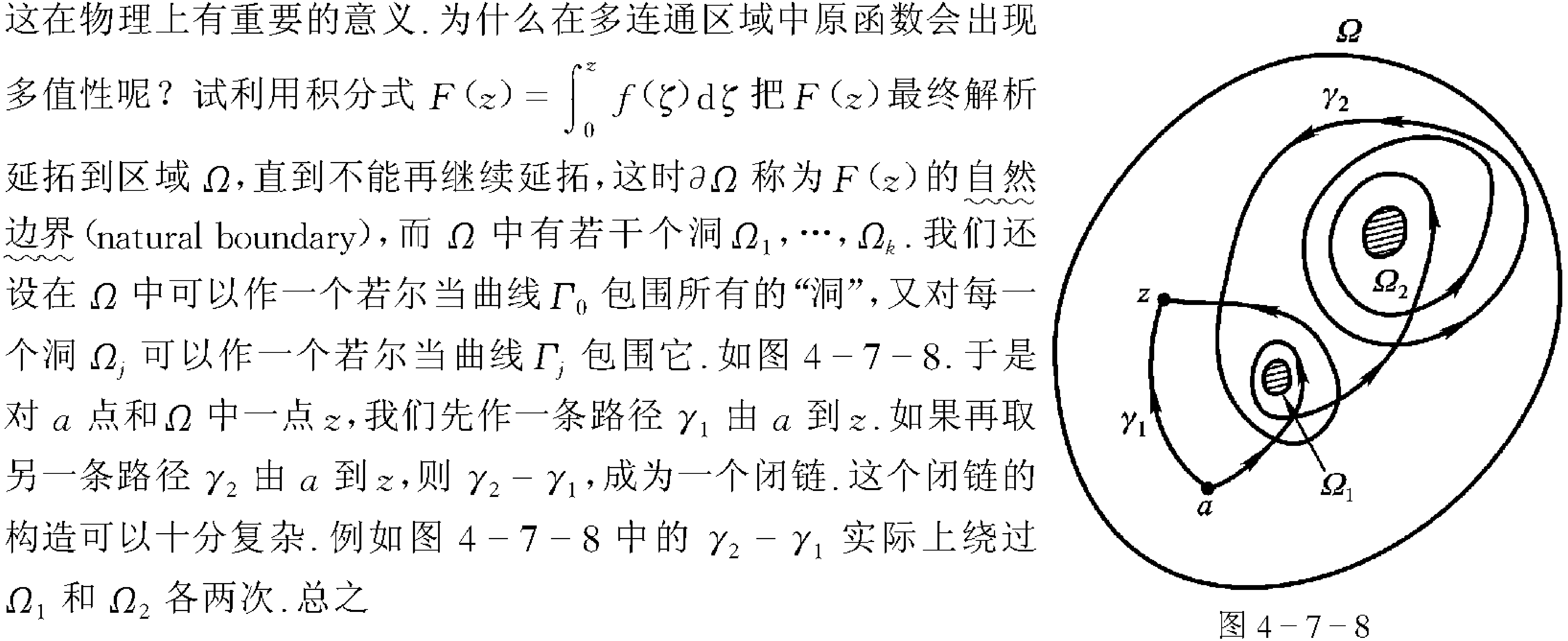



图 4－7－8

$$\gamma_2-\gamma_1\sim\sum_{j=1}^{n}c_j\Gamma_j.$$

这里 $c_j$ 是正负整数或 0. 而由柯西定理的同调论表述(定理 4)有

$$\int_{\gamma_2-\gamma_1}f(\zeta)\mathrm{d}\zeta=\int_{\sum_j c_j\Gamma_j}f(\zeta)\mathrm{d}\zeta=\sum_j c_j\int_{\Gamma_j}f(\zeta)\mathrm{d}\zeta$$
$$=\sum_j c_j\theta_j,$$

这里 $\theta_j=\int_{\Gamma_j}f(\zeta)\mathrm{d}\zeta$，而 $c_j$ 是整数.

$$\int_{a(\gamma_2)}^{z}f(\zeta)\mathrm{d}\zeta=\int_{a(\gamma_1)}^{z}f(\zeta)\mathrm{d}z+\sum_j c_j\theta_j.$$

这个式子明确地告诉我们，多值的原函数之任意值一定是它的某一指定值再加上 $\theta_j$ 的“整系数线性组合” $\sum_j c_j\theta_j$. 这些 $\theta_j$ 称为“周期”. 其实我们在第二章中就已见到

$$\mathrm{Ln}\,z=\ln z+k\cdot(2\pi\mathrm{i}),$$

这里的 $k$ 就是整系数 $c_j$，$2\pi\mathrm{i}$ 就是 $\theta_j$. 不过我们看到其实周期是说的对数函数的反函数 $z=\mathrm{e}^w$ 有周期 $2\pi\mathrm{i}$.

细心的读者马上就会看到，上面的讨论很不严格. 怎样证明 $\gamma_2-\gamma_1$ 和 $\sum_j c_j\Gamma_j$ 同调呢？这就涉及对有洞的区域上的闭链之集合如何刻画的问题. 由此引进了同调群这个代数拓扑学中的

中心概念.可见,一旦把积分引入复域而考虑解析函数的积分时,这些拓扑学的问题就很自然地浮出水面了.尤其令人吃惊的是,在 20 世纪后半期,发现了它与物理学有十分密切的关系.

**3. 柯西积分公式** 我们不但发现了解析函数的积分有如此丰富的性质,进一步还发现了,所有的单值解析函数又一定都可以表示为一种特殊类型的积分.它不但与上面讲的拓扑学问题有关,而且还与上一节讲的 $\delta$ 函数有关.

下面我们介绍解析函数的这种积分表示法——柯西积分公式.它是研究解析函数性质的重要工具,但它的应用范围并不只限于复变量的解析函数,而且对实变量函数的研究也有很大的意义.所以,我们将在实的情况下来讲述其基本定理,然后再把它用于复的情况.但为此,我们先用复记号表示实函数.例如 $\Omega\subset\mathbf{C}$ 就看成 $\mathbf{R}^2$ 中一个区域,这时我们用另一个字母 $\omega$ 来表示它.因此,下面凡说到 $C^1(\bar{\omega})$就指实变量$(x,y)$在 $\bar{\omega}$上的复值 $C^1$ 函数,而非 $z$ 的解析函数,而反之 $C^1(\bar{\Omega})$则约定表示复变量的连续可微函数,因此一定是解析的.同样,对 $u\in C^1(\bar{\omega})$,$\dfrac{\partial u}{\partial \bar{z}}=\dfrac{1}{2}\left(\dfrac{\partial u}{\partial x}+\mathrm{i}\dfrac{\partial u}{\partial y}\right)$只是一个形式记号.

**定理 9(柯西积分公式)** 若 $u\in C^1(\bar{\omega})$,则对 $z\in\omega$,

$$u(\zeta)=(2\pi\mathrm{i})^{-1}\int_{\partial\omega}\frac{u(z)}{z-\zeta}\mathrm{d}z+(2\pi\mathrm{i})^{-1}\iint_{\omega}\frac{\partial u/\partial\bar{z}}{z-\zeta}\mathrm{d}z\wedge\mathrm{d}\bar{z}. \tag{18}$$

**证** 记 $\omega_\varepsilon=\{z;z\in\omega,|z-\zeta|>\varepsilon\}$,这里 $\varepsilon$ 是充分小正数.对 $u(z)/(z-\zeta)$应用格林公式(17),注意$\partial/\partial\bar{z}$虽然是一个形式记号,但莱布尼茨乘积求导公式仍适用于它,这一点读者容易自行验证.又因$\dfrac{1}{z-\zeta}$当$\zeta\neq z$ 时对 $z$ 是解析的,所以在 $\omega_\varepsilon$ 中$\dfrac{\partial}{\partial\bar{z}}\left(\dfrac{1}{z-\zeta}\right)=0$.这样,由(17)注意到 $\partial\omega_\varepsilon=\partial\omega-\{z:|z-\zeta|=\varepsilon\}$,即得

$$\iint_{\omega_\varepsilon}\frac{\partial u/\partial\bar{z}}{z-\zeta}\mathrm{d}\bar{z}\wedge\mathrm{d}z=\int_{\partial\omega}u(z)(z-\zeta)^{-1}\mathrm{d}z-\mathrm{i}\int_0^{2\pi}u(\zeta+\varepsilon\mathrm{e}^{\mathrm{i}\theta})\mathrm{d}\theta.$$

这里我们应用了在$|z-\zeta|=\varepsilon$ 上$z-\zeta=\varepsilon\mathrm{e}^{\mathrm{i}\theta}$,而$|z-\zeta|=\varepsilon$ 的正向是$\theta$ 由 0 增加到 $2\pi$.令 $\varepsilon\to0$ 即得(18)式.

(18)式无论对实的情况与复的情况都极重要,而我们现在要把它应用于复情况:特别限于 $u$ 在 $\bar{\Omega}$上解析而且$\partial$

个特定概念来刻画它.

**定理 11** 若 $L$ 为一分段 $C^1$ 闭曲线,而 $a\notin L$,则

$$n(L,a)=\frac{1}{2\pi i}\int_L\frac{dz}{z-a} \tag{20}$$

必为一整数,称为 $L$ 对于 $a$ 的环绕数(winding number).

**证** $\frac{1}{z-a}$作为 $z$ 的函数其原函数为 $\ln(z-a)$,故

$$n(L,a)=\frac{1}{2\pi i}\ln|z-a|\Big|_L+\frac{1}{2\pi}\arg(z-a)\Big|_L,$$

这里 $f(z)|_L$ 表示 $f(z)$在 $L$ 上的变幅:任取 $z_0\in L$,$f(z)|_L$ 表示 $f(z)$从 $f(z_0)$开始,再令 $z$ 沿 $L$ 解析延拓一周回到 $z_0$ 后的终值之增加值.记 $z-a=e^{i\varepsilon}$,前一项为$\frac{1}{2\pi i}\ln r|_L=0$,后一项为 $\theta|_L$,即是 $z-a$ 的辐角当 $z$ 依正向绕 $L$ 一周(即 $z$ 沿 $L$ 正向绕 $a$ 回到起点)的增值.所以$\frac{1}{2\pi}\theta\Big|_L$ 即环绕的次数.

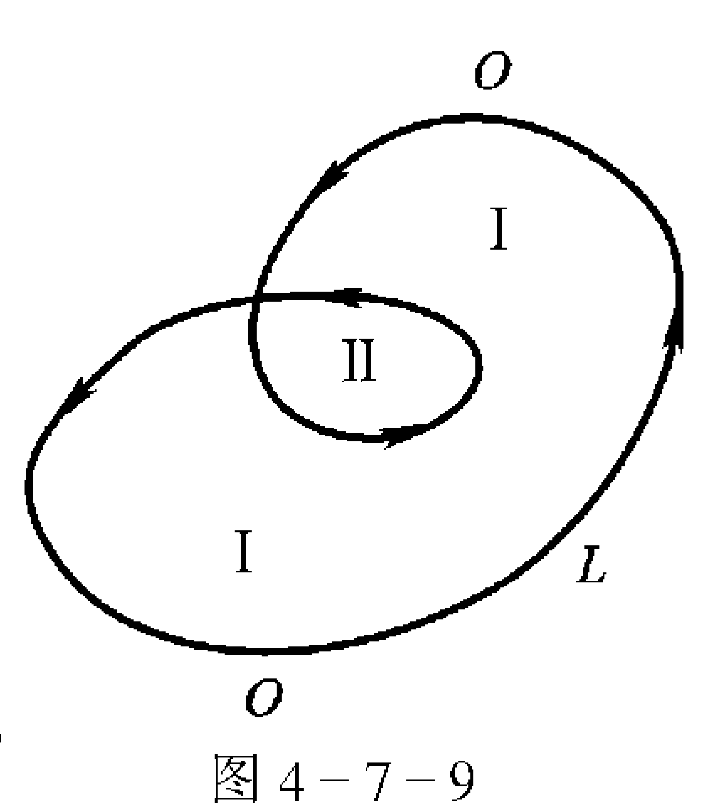


图 4-7-9

这个定理虽然简单,却有丰富的含意.首先,它告诉我们,在前面定理 5~8 中我们一再强调原函数从局部观点看与从整体观点看是不同的:这里正说明了$\frac{1}{z-a}$的原函数 $\ln(z-a)$从整体看是多值的.其次,正是这种多值性给我们提供了一个刻画 $L$ 的拓扑性质的重要的不变量——环绕数,它在拓扑学中有重要的地位.本节一开始就介绍了著名的若尔当定理,但未给证明.不证明并不等于说它不重要,而是由于证明太长,而且这定理本身未必能很明晰地说明问题.如果路径比图 4-7-8 还复杂,怎么能确定哪里是内,哪里是外?给了一条闭曲线 $L$ 就把复平面划分为若干块,但不一定是两块,图 4-7-9 就是三块.这个图形所表示的正是一个复形,它由两部分组成,一是整个最外面的心脏形曲线所包围的 $\mathbf{C}$ 平面上的区域,可以写成 Ⅰ + Ⅱ,另一部分则是内部的曲线所围成的,它也是Ⅱ,但却不能与 $\mathbf{C}$ 平面上的 Ⅰ + Ⅱ 混淆,否则按图上所给的正向来分别区域的内、外就会有困难,所以把它看成一个复形(Ⅰ + Ⅱ) + Ⅱ = Ⅰ +2Ⅱ 最好.还有一块即含无穷远点的无界区域 $O$.由定理 3,仍有 $\int_L f(z)dz=0$. $O$,Ⅰ 和Ⅱ 三块之中一定有一块即我们这里的 $O$ 包含无穷远点,不妨称为外域(如果 $L$ 本身就通过无穷远点,则要修正我们的讨论.在复分析中怎样讨论无穷远点有一套办法,并在理论上有充分根据,我们不能在这里讲了).我们马上看出环绕数以下的性质.

(i) 若 $L$ 是不自交的若尔当曲线(图上的 $L$ 是自交的因此不是若尔当曲线),则当 $a$ 在外域中时 $n(L,a)=0$,$a$ 不在外域中(即在内域中)时 $n(L,a)=1$.

(ii) 若 $L$ 是允许自交的闭曲线,则虽然仍有外域,内域是什么就难说了.图 4-7-9 中的Ⅰ,Ⅱ哪一个在哪一个之内呢?为了区别各块,我们把 $n(L,a)$看成 $a\in\mathbf{C}\setminus L$ 的函数,容易看到

$$\frac{d}{da}n(L,a)=\frac{1}{2\pi i}\int_L\frac{dz}{(z-a)^2}=-\frac{1}{2\pi i}\int_L\frac{d}{dz}\left(\frac{1}{z-a}\right)dz.$$

这时被积函数又有原函数 $-\frac{1}{z-a}$,但它是单值函数,而与$\frac{1}{z-a}$的原函数 $\ln(z-a)$为多值函数

不相同.因此，$\frac{\mathrm{d}}{\mathrm{d}a}\ln(L,a)=0$，从而 $n(L,a)=\text{const}$，只要 $a$ 在变化中不触及 $L$ 即可.就是说 $n(L,a)$ 是一个分块常值函数.例如在图 4－7－9 中

$$n(L,a)=\begin{cases}0, & a\in 0;\\ 1, & a\in \text{Ⅰ};\\ 2, & a\in \text{Ⅱ}.\end{cases}\tag{21}$$

当然也可能有这样的 $L$，对于位于不同"块"中的 $a$，$n(L,a)$ 取相同值，但是不论如何，现在用一个可计算的积分来刻画 $L$ 的几何性质，比之若尔当定理，总是更容易理解和捉摸.

现在再回过头来看柯西积分公式(19).那里我们假设了 $\partial\Omega$ 是一可求长若尔当曲线，因此对何处是其内域，何处是其外域是很清楚的.但是如果考虑 $\frac{1}{2\pi\mathrm{i}}\int_L\frac{f(\zeta)\mathrm{d}\zeta}{\zeta-z}$，而 $L$ 是一个一般的闭曲线，即可以自交，可以缠绕 $\Omega$ 的某一个洞多次，等等，如图 4－7－8 那样，柯西公式应如何改写？从(19)看，如果令记 $\partial\Omega$ 为 $L$，则因 $L$ 把平面分成两部分，即 $\Omega$ 与其外域，很容易看到 $L$ 的环绕数是

$$n(L,z)=\begin{cases}0, & z\text{ 在 }\Omega\text{ 外},\\ 1, & z\text{ 在 }\Omega\text{ 内}.\end{cases}$$

所以，(19)可以写为

$$\frac{1}{2\pi\mathrm{i}}\int_L\frac{f(\zeta)\mathrm{d}\zeta}{\zeta-z}=n(L,z)f(z).\tag{22}$$

于是我们会想到，若 $L$ 为区域 $\bar{\Omega}$ 中之闭链，而且 $L\sim 0$，则对在 $\Omega$ 中解析，在 $\Omega\cup L$ 上连续的 $f(z)$(如果 $L\subset\Omega$，则 $f(z)$ 在 $\Omega\cup L$ 中连续的条件自动成立)，应有(22)式.它就是一般的柯西积分公式.这当然是要证明的，而为此就得把同调的概念会仔细讲清楚，我们这里就略去不讲了.

**注 1**　以上的讨论全是以对数函数的多值性为基础的，而对数函数 $\ln(z-\alpha)$ 则是 $\frac{1}{z-\alpha}$ 的原函数：$\ln(z-\alpha)=\int_a^z\frac{\mathrm{d}z}{z-\alpha}+C$.因此在许多较新的复分析的教材中就直接以 $\int_1^z\frac{\mathrm{d}z}{z}$ 作为对数函数 $w=\ln z$ 之定义.这样做最大的优点在于立即看出环绕数与对数的关系，但是对数函数的其它性质则不那么清楚.更重要的是这种作法的优点在实变量的情况下看不出来.注意到这个原函数 $w=\int_1^z\frac{\mathrm{d}z}{z}$ 必定是以下柯西问题之解：

$$\frac{\mathrm{d}w}{\mathrm{d}z}=\frac{1}{z},$$
$$w(1)=0,$$

我们也可以用上述常微分方程柯西问题之解作为对数函数的定义.尤其是，这个函数的反函数 $z=z(w)$ 是下面的柯西问题之解

$$\frac{\mathrm{d}z}{\mathrm{d}w}=z,$$
$$z(0)=1,$$

而可以用这个常微分方程柯西问题之解作为指数函数的定义.我们在第二章中就是这样做的.本来，一个函数常有多种不同的定义方式.采用哪一个方式全看我们想要研究的是此函数的哪些性

质;用哪一种定义方式又能更快地达到目的,并能更好地对这些性质获得新的洞察,发现新的联系.正是从这一点出发,我们在第二章中对一些最常见的初等函数都采用了微分方程来定义它们.

还要提到一点,在上一节我们指出一个具有第一类间断点的函数,其 $D'$ 导数恰好是 $\delta(x)$,即

$$H(x)=\int_{-\infty}^{x}\delta(t)\mathrm{d}t$$

现在则有(我们把 $a$ 记作 $x$,尽管它仍是复的)

$$n(L,x)=\frac{1}{2\pi\mathrm{i}}\int_L\frac{\mathrm{d}z}{z-x}$$

可见 $\dfrac{1}{2\pi\mathrm{i}(z-x)}$ 多少与 $\delta$ 函数有些类似.再说

$$I(x)=\frac{1}{2\pi\mathrm{i}}\int_L\frac{f(\zeta)}{\zeta-x}\mathrm{d}z=\begin{cases}f(x), & x\text{ 在 }L\text{ 内},\\ 0 & x\text{ 在 }L\text{ 外}.\end{cases}$$

所以 $I(x)$ 之内、外极限之差恰为"密度"函数 $f(x)|_{x\in L}$,这与 $\delta$ 函数之"再生性"(reproducing property)

$$f(x)=\int_{-\infty}^{\infty}f(\zeta)\delta(\zeta-x)\mathrm{d}\zeta$$

也十分相似.怎样从广义函数观点去看待柯西积分公式是一个很诱人的问题,而且在理论物理——量子场论的发展中起了作用.

把微积分学引入复域我们在本书中是分成两部分进行的.在第三章中先把微分学引入复域,并且得到了关于解析函数的两个定义.一个定义说所谓的解析函数 $f(z)$ 就是在一点的某个邻域中连续可微的函数.由此得到 C-R 方程.而且从几何上看,$w=f(z)$ 是由 $z$ 平面到 $w$ 平面的一个映射,且在 $f'(z)\neq 0$ 处,这个映射保持角度不变,因此称为共形映射.这种几何的观点是黎曼的观点.我们不妨称这样定义的解析函数为 C-R 意义下的解析函数.第二种定义的方式是说所谓 $f(z)$ 在 $z$ 点解析就是在 $a$ 的某个圆形邻域 $|z-a|<r$ 中可以展开为幂级数的函数.用幂级数去研究解析函数是魏尔斯特拉斯的主要方法和贡献.由幂级数的性质知道,在收敛圆内 $f(z)=\sum\limits_{n=0}^{\infty}a_n(z-a)^n$ 一定可以无限阶求导(自然也是连续求导),所以这种解析函数必为 C-R 意义下的解析函数.那么,其逆是否成立?好在我们还有第三个途径.若 $f(z)$ 是 C-R 意义下的解析函数,利用格林定理立即可得关于积分的柯西定理(当然也就适用柯西积分公式).但柯西定理之逆摩列拉定理也成立.所以我们还可以说.适合柯西定理的函数的就是解析函数.这样一来,解析函数就有了三种定义.其关系如下:

幂级数意义(或称魏尔斯特拉斯意义)⇒C-R 意义⇒柯西定理意义.

所以我们只要能再证明:柯西定理意义⇒幂级数意义,即证明:若 $f(z)$ 在 $\Omega$ 中适合柯西定理,则 $f(z)$ 处可在 $\Omega$ 内任一点 $a$ 的邻域中展开为幂级数,就得知解析函数的这三种定义是等价的.特别是 C-R 意义的与魏尔斯特拉斯的解析函数是一样的.

当然,上面讲的都是局部意义下的解析函数.

现在我们来证明,若 $f(z)$ 在 $a$ 点附近连续可微(从而满足 C-R 方程,从而是 C-R 意义下的解析函数),必在这点附近任意阶连续可微.实际上,满足 C-R 方程,则亦满足柯西积分公式.在 $a$ 点的这一个"附近"中取开区域 $\Omega$,使 $\partial\Omega$ 为分段 $C^1$ 曲线 $L=\partial\Omega$,则对 $z\in\Omega$ 有

$$f(z)=\frac{1}{2\pi i}\int_{\partial\Omega}\frac{f(\zeta)}{\zeta-z}d\zeta. \tag{23}$$

因为 $z\in\Omega$ 故 $z\notin\partial\Omega$ 而上式中的分母 $\zeta-z\neq0$. 用含参变量 $z$ 的积分对 $z$ 之可微性（$\frac{1}{\zeta-z}$对 $z$ 之可微性在第三章已证明了），即知由(23)定义的 $f(z)$在 $\Omega$ 中连续可微，其导数是

$$f'(z)=\frac{1}{2\pi i}\int_{\partial\Omega}\frac{f(\zeta)}{(\zeta-z)^2}d\zeta,$$

再用一次含参变量积分的知识知 $f'(z)$在 $\Omega$ 中连续可微.（我们也可以分开被积函数的实、虚部再用实黎曼积分的知识，证明 $f'(z)$之实、虚部的偏导数适合 C－R 方程. 这一切都不再说了.）类似地有

$$f^{(n)}(z)=\frac{n!}{2\pi i}\int_{\partial\Omega}\frac{f(\zeta)}{(\zeta-z)^{n+1}}d\zeta. \tag{24}$$

所以，若 $f(z)$在 $\Omega$ 中连续可微，它必在其中任意阶连续可微.

现取 $a\in\Omega$ 而作 $f(z)$在 $a$ 点的形式泰勒级数

$$\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(z-a)^n=\frac{1}{2\pi i}\sum_{n=0}^{\infty}\int_{\partial\Omega}f(\zeta)\frac{(z-a)^n}{(\zeta-a)^{n+1}}dz \tag{25}$$

并研究它的收敛性，因为 $a\notin\partial\Omega$，故由 $a$ 到$\partial\Omega$ 之距离 $r>0$，在圆$|z-a|\leqslant r_0<r$ 中任取一点 $z$，则

$$|z-a|\leqslant r_0<r\leqslant|\zeta-a|,\quad \zeta\in\partial\Omega,$$

所以$\left|\frac{z-a}{\zeta-a}\right|\leqslant\frac{r_0}{r}=\theta<1$，从而级数

$$\sum_{n=0}^{\infty}\frac{(z-a)^n}{(\zeta-a)^{n+1}}=\frac{1}{\zeta-a}\sum_{n=0}^{\infty}\left(\frac{z-a}{\zeta-a}\right)^n$$

$$=\frac{1}{\zeta-a}\frac{1}{1-\frac{z-a}{\zeta-a}}=\frac{1}{\zeta-z}$$

在$|z-a|\leqslant r_0$，$\zeta\in\partial\Omega$ 上对 $\zeta$ 一致收敛. 这样可以用逐项积分求得(25)之值，而有

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(z-a)^n&=\frac{1}{2\pi i}\int_{\partial\Omega}f(\zeta)\sum_{n=0}^{\infty}\frac{(z-a)^n}{(\zeta-a)^{n+1}}d\zeta\\&=\frac{1}{2\pi i}\int_{\partial\Omega}\frac{f(\zeta)}{\zeta-z}dz=f(z),\end{aligned}$$

这个式子在$|z-a|\leqslant r_0$ 中成立. 但是我们不能说 $r_0$ 或 $r$ 就是泰勒级数的收敛半径$R$，后者是$\left[\overline{\lim_{n\to\infty}}\sqrt[n]{|f^{(n)}(a)|/n!}\right]^{-1}$，是由 $f(z)$在 $a$ 点"附近"的状况决定的. 现在我们的 $\Omega$ 选得相当任意，因此上面的 $r_0$ 和 $r$ 也相当任意. 但是我们决不可能取一个把$|z-a|<R$ 包含在内的区域选作$\Omega$，不然的话，我们将会得出级数(25)在收敛圆$|z-a|<R$ 外仍收敛的结论，而这与收敛半径的定义是矛盾的. 这样我们就看出，任一解析函数在任一点 $a$ 的泰勒级数的收敛圆周上必有其一个"奇点"，要不然的话，就可以取上述更大的 $\Omega$ 了.

至此，我们已完全明白了，C－R 意义下的解析函数与魏尔斯特拉斯意义下的解析函数是完全一致的.

# 第五章　傅里叶级数与傅里叶积分

## §1　傅里叶级数——从什么是谱谈起

**1. 什么是谱**　傅里叶级数和傅里叶积分理论系统地首次出现在傅里叶(Jean-Baptiste Joseph Fourier，1768—1830)的名著《热的解析理论》(Théorie Analytique de Chaleur，有中译本，武汉出版社，1993)中. 这是一部伟大著作. 可以毫无疑问地说，它开创了19世纪数学和物理学的新篇章. 麦克斯韦称它是“一首伟大的数学诗”. 将近两百年来它的理论和方法渗透到几乎一切数学物理分支中，成为一个十分广泛的数学分支——调和分析，可以说长盛不衰. 其实傅里叶这本书最开始是为了应征巴黎科学院的悬赏求解而在1807年提交的一篇报告. 到1822年正式成书. 书上开宗明义就指出，自然哲学(当时人们对数理科学的称谓)的目的就在于探讨自然界的规律性. 牛顿和他的先驱们虽然已取得伟大成就，星体的运动、光的本性、声的传播等等都被归结为几个最基本的规律，但是热却未探讨过. 而热现象又是宇宙中最常见、最基本的物理现象. 傅里叶研究它，不只着重于国计民生的实际应用(当然他丝毫也未轻视这一切)，而且还有例如下述的种种问题：地球处于太阳的热辐射中，既从太阳吸热，又把热量辐射到空中，怎样才能达到平衡？因为太阳的热辐射而产生的大气运动，怎样影响气候？同样因太阳的热辐射而产生的洋流有什么规律性？傅里叶认为这一切都是牛顿未曾涉足的领域. 可是在这部书中同样地没有回答这些问题，而且他关于热的基本概念是错误的，这一点我们在第三章§1中讲过. 那么傅里叶的功绩何在呢？他在这本书中引用牛顿在《自然哲学的数学原理》序言中一段话：“几何学的荣耀在于，它从别处借用很少的原理，就能产生如此众多的成就.”这其实也是牛顿的方法论：把一切机械运动归结为几个根本定律，并进而写成一个二阶常微分方程组：$m_i\dfrac{\mathrm{d}^2x_i}{\mathrm{d}t^2}=f_i(t,x,\dot{x})$. 傅里叶也力图采取这个方法论，他指出关于热的纷纭万象，其实都可归结为热的贮存(比热)与传导(导热率)，然后明确给出一个二阶偏微分方程——热传导方程. 这个方程我们已在第三章§1中详细讲过. 其次他为求解这个偏微分方程提出了系统的方法. 这个方法的要点我们将在下面讨论. 现在的问题是，经过将近200年的发展，傅里叶的理论与方法已得到极大的发展，我们可否回顾一下，从中得到一些有助于了解人们认识自然界的富有成果的思想？从谱论的角度来看待傅里叶的理论与方法(以下简称为傅里叶分析或调和分析)无疑是十分有教益的.

那么什么是“谱”呢？19世纪关于光谱的研究是物理学的最重要的问题之一. 光谱学(spectroscopy)是迈克尔孙(Michaelson)和莫莱(Morley)创立的. 当时他们测量了镉的谱线，并以其波

长作为长度的标准. 迈克尔孙因此获得 1907 年的诺贝尔物理学奖. 这个长度标准一直使用到 20 世纪 60 年代才使用铯 133 的一条谱线代替它. 光谱学被用来从复杂的信息中获取有用的信息,或者在如有人说的,去寻找“隐藏着的周期性”. 这是怎么回事呢?让我们用一个历史早得多因而更容易理解的力学模型来说明.

第三章中介绍的偏微分方程,除了热传导方程以外,都出现于 18 世纪中期以前,都是牛顿力学对某些具体情况如弹性体、流体上应用的结果. 丹尼尔·伯努利(Daniel Bernoulli)研究了弦的横振动. 他的基本模型首先是把弦看成是由微小的物体——微元——组成的因而可以应用质点组的牛顿力学. 不过这些质点互相之间是有联系的,这种联系就表现为弦的张力. 因而,一个质点的振动就带动了其左右其它质点振动. 振动于是沿弦传播开去,就成了波. 我们把这个思想用于杆的纵振动. 虽然从几何上看不如弦的横振动那么直观,但其物理机制更容易理解. 于是设有一杆,位于 $x$ 轴上区间 $[0,L]$ 处. 我们认为杆是位于等分点上的 $N+1$ 个质点组成的,其质量都相同为 $m$,间距也都相同为 $a$,而且相邻的质点是由相同的弹簧连接起来的. 令 $q_i(t)$ 表示第 $i$ 个质点之位移(从 $t=0$ 算起). $i=0$ 与 $i=N$ 则是杆的两个端点. 设杆的两端是固定的,即 $q_0(t)=q_N(t)=0$,现在我们要求其它的 $q_i(t)$,这里 $i=1,2,\cdots,N-1$. 第 $n$ 个质点原来位于 $x=x_n=\dfrac{nL}{N}$ 处,在时刻 $t$,因为它有了位移 $q(x_n,t)$,所以位于 $x_n+q(x_n,t)$ 处. 同样第 $n-1$ 和 $n+1$ 个质点现在分别位于 $x_{n-1}+q(x_{n-1},t)=(x_n-a)+q(x_n-a,t)$ 以及 $x_{n+1}+q(x_{n+1},t)=(x_n+a)+q(x_n+a,t)$ 处. 于是例如弹簧 $[x_n,x_{n+1}]$ 之长成为 $(x_n+a)+q(x_n+a,t)-x_n-q(x_n,t)=a+q(x_n+a,t)-q(x_n,t)$,其相对伸长(即长度的增量与原来长度之比,亦即弹性力学中的形变)为 $\dfrac{q(x_n+a,t)-q(x_n,t)}{a}$. 同理,弹簧 $[x_{n-1},x_n]$ 的形变为 $\dfrac{q(x_n-a,t)-q(x_n)}{a}$. 弹性力学告诉我们,$x_n$ 处的质点受的力与形变成正比,其比例系数我们假设对各个弹簧都是一样的,为 $K$,因此,第 $n$ 个质点的运动方程为

$$m\ddot{q}_n=\frac{K}{a}(q_{n+1}-2q_n+q_{n-1}).$$

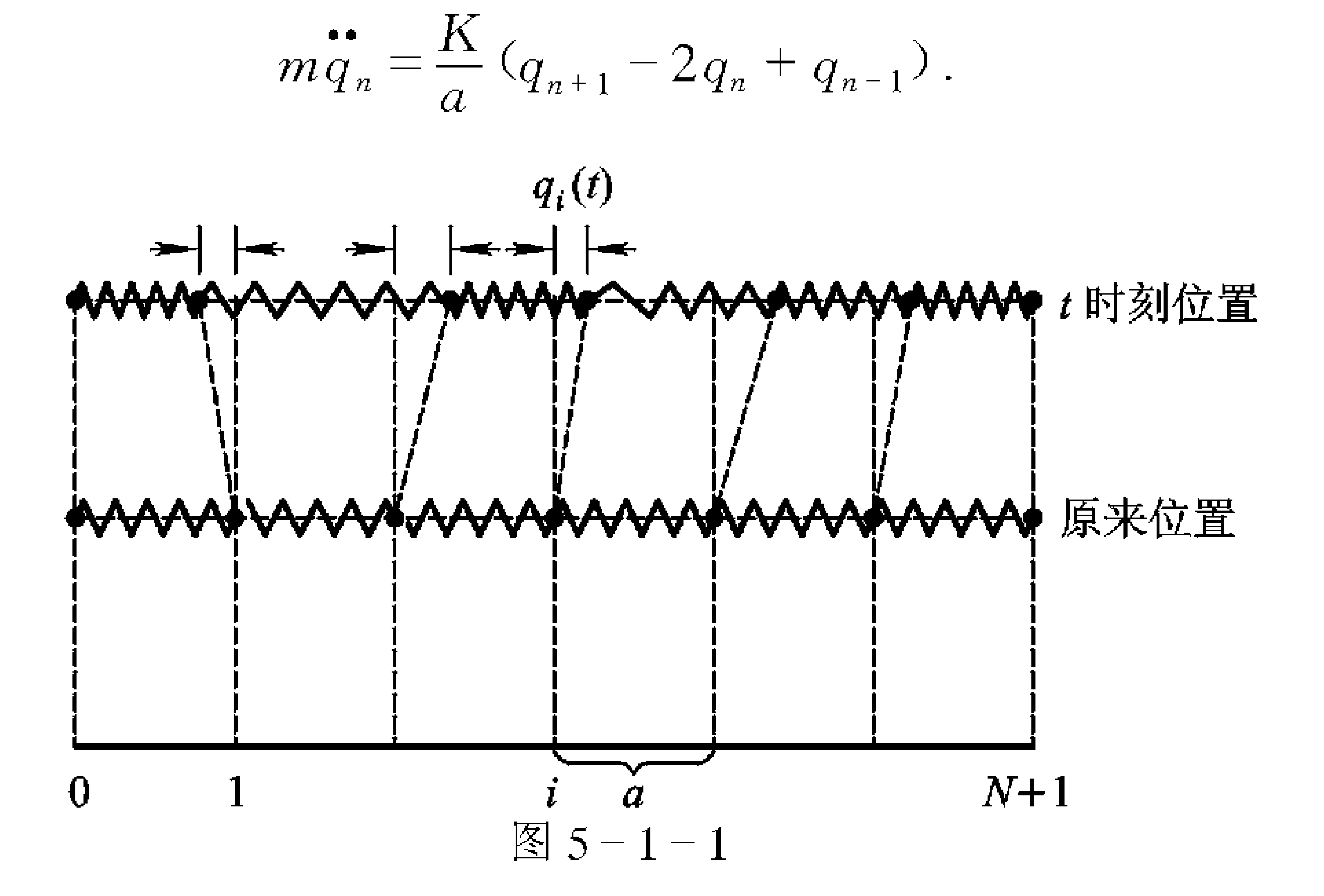


图 5-1-1

如果杆的质量是平均分布在杆上的,则第 $n$ 个质点之质量可写为 $m=\rho a$,$\rho$ 是密度,它也是一个常数,于是我们有方程组

$$\rho\ddot{q}_n=c(q_{n+1}-2q_n+q_{n-1}).\tag{1}$$

以上 $c=K/a^2$，$q_n=q(x_n,t)$.

(1) 是一个二阶常微分方程组，我们可以把它写成矩阵形式

$$\rho\ddot{q}=cAq. \tag{2}$$

这里 $q$ 是一个列向量 ${}^t(q_0,\cdots,q_N)$，除了 (2) 以外，本来我们还有两个"边值条件"

$$q_0=q(0,t)=0,\ q_N=q(l,t)=0, \tag{3}$$

但是为了讨论问题方便起见，我们不妨用 $q_0(t)=q_N(t)$ 来代替 (3)，这就好比我们在考虑一个首尾相连的环形杆，虽然不自然，却能更好地说明我们的方法的实质. 矩阵 $A$ 是一个三对角矩阵

$$A=\begin{pmatrix} -2 & 1 & & & & & \\ 1 & -2 & 1 & & & & 0 \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & 0 & & & 1 & -2 & 1 \\ & & & & & 1 & -2 \end{pmatrix} \tag{4}$$

为了求解这个方程组，从数学上说，我们应该设法把未知向量换成一个新的未知向量 $a(t)={}^t(a_0(t),\cdots,a_N(t))$，而使系数矩阵成为一个对角矩阵. 现在，我们把记号整理一下：设连同两端一共有 $N+1$ 个点，重新编号为 $a_0(t),\cdots,a_N(t)$，0 和 $N$ 是两个端点，而边值条件是 $a_0(t)=a_N(t)$. 这样，我们需求的只有 $a_1(t),\cdots,a_N(t)$.

把 $q(t)$ 换成 $a(t)$ 就相当于作线性变换

$$q_n(t)=\sum_k a_k(t)u_n^k. \tag{5}$$

注意，我们这里重新编号，于是 $n$ 和 $k$ 都从 1 变到 $N$. 这在物理上是什么意义？常微分方程组 (1) 是一个"耦合"的 (coupled) 方程组，意即第 $i$ 个质点的运动受到第 $i-1$ 和第 $i+1$ 两个质点的"牵扯"，而且仿此，所有 $N$ 个质点都牵扯在一起. 把 $A$ 化成对角形，得到一个形如

$$\ddot{a}=\frac{c}{\rho}\mathscr{A}a$$

的常微分方程组，其未知函数向量是 $a={}^t(a_1(t),\cdots,a_N(t))$，$\mathscr{A}$ 是一个对角矩阵

$$\mathscr{A}=\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix},$$

而方程组成为

$$\ddot{a}=\frac{c}{\rho}\lambda_i a_i, \quad i=1,\cdots,N. \tag{6}$$

所有的 $a_i$ 都互不牵扯，物理上称为"解耦"(uncoupled). 一旦实现了解耦，解方程自然很容易. 可是在物理上重要的是这个新向量 $a$ 有特殊的物理意义，而反映了这个物理系统的物理本质. 所以在物理上，这个新的未知函数向量称为简正模式 (normal modes)，其各个分量称为简正坐标 (normal coordinates)：即原来的位置向量 $q(t)$ 对于基函数 (basis functions) $u_n^k$ 的线性展开式之系

数(对 $k$ 求和, $n$ 则是 $q(t)$之分量的标号),至于在数学上它们是什么,下面可以看得很清楚.

下面我们来研究如何求 $u_n^k$ 和 $a(t)$. 我们已经说了,为方便起见我们不妨讨论周期边值条件 $q_o(t)=q_N(t)$. 我们甚至不妨认为 $q_n(t)$对一切 $n$ 都有定义,而且适合 $q_n(t)=q_{N+n}(t)$. 这就相当于把定义在区间$[0,l]$上的函数周期地拓展到整个 $\mathbf{R}$ 上. 既然如此,由(5)式可见 $u_n^k$ 对于 $n$ 也应有周期性 $u_n^k=u_{n+N}^k$,为了适合这个要求,我们取

$$u_n^k=\frac{1}{\sqrt{N}}\mathrm{e}^{\mathrm{i}k\alpha n} \tag{7}$$

而 $\mathrm{e}^{\mathrm{i}k\alpha N}=1$,所以我们令

$$k\alpha=\frac{2\pi l}{N},\quad l\text{ 为整数}, \tag{8}$$

$\alpha=L/N$(就是前述两质点的距离 $a$),所以有 $k=2\pi l/L$,至于(6)式中的因子$\frac{1}{\sqrt{N}}$是为了规范化(物理学家喜欢说“归一化”,其实都是为了把某个向量变成单位长向量)(8)式中的整数 $l$ 限取 $0<l\leqslant N$ 值,这是因为 $\mathrm{e}^{\mathrm{i}k\alpha n}=\mathrm{e}^{\mathrm{i}2\pi ln/N}$对 $l$ 有周期$N$,当 $l$ 越出这个范围时同样的 $u_n^k$ 将会重现. 这就相当于物理上说的“分辨力”不够.

现在再来考查(7)式, $n\alpha$ 实际上代表位置,所以是一个空间变量,它与 $\mathrm{e}^{\mathrm{i}kx}$ 相近,而后者是平面波, $k$ 称为波数,其量纲为$[L]^{-1}$. (7)中的 $k$ 也是一样,所以(7)实际上是正弦形状的平面波的离散化,它的重要性质是

**定理 1**　矩阵 $U=(u_n^k)$($k$ 表示行, $n$ 表示列)是一个酉矩阵.

**证**　我们只要证明以下两个等式即可:

$$\sum_{n=1}^{N}\bar{u}_n^{k'}\cdot u_n^k=\delta_{kk'}, \tag{9}$$

$$\sum_{k=1}^{N}\bar{u}_{n'}^{k}\cdot u_n^k=\delta_{nn'}. \tag{10}$$

其实它们只是作简单的三角运算(由简单的三角学公式可以得到如此重要的定理实在大为值得注意). 为证(9)式,令 $k'=\frac{2\pi l'}{Na}$, $k=\frac{2\pi l}{Na}$,则 $l\neq l'$时

$$\begin{aligned}\sum_{n=1}^{N}\bar{u}_n^{k'}\cdot u_n^k&=\frac{1}{N}\sum_{n=1}^{N}\mathrm{e}^{2\pi\mathrm{i}(l-l')n/N}\\&=\frac{\mathrm{e}^{2\pi\mathrm{i}(l-l')/N}}{N}\cdot\frac{1-\mathrm{e}^{2\pi\mathrm{i}(l-l')}}{1-\mathrm{e}^{2\pi\mathrm{i}(l-l')/N}}=0.\end{aligned}$$

这里我们应用了等比数列的求和公式,因为 $l$ 与 $l'$均在 1 到 $N$ 中, $l-l'$不可能是 $N$ 的整数倍,因而分母不能为 0,但分子一定为 0.

当 $l=l'$时(9)式自然成立.

(10)式的证明与此类似.

还要注意到, $\bar{u}_n^k=u_n^{-k}$,而 $q(t)$应该是实值的,因此(5)式中应取 $\bar{a}_k(t)=a_{-k}(t)$. 注意到这样一些事实以后,我们就完全转到复域中去考查方程组(2). 这里 $A$ 本来是实对称方阵. 现在看作复的埃尔米特方阵. 线性代数知识告诉我们,它一定可以用一个酉矩阵 $U$ 对角化:

$$U^{-1}AU=\mathscr{A}=\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_N\end{pmatrix}.$$

$\lambda_1,\cdots,\lambda_N$ 是 $A$ 的特征根.故若将向量 $q$ 变化如下：

$$q(t)=Ua(t),$$

方程(2)就成为(6)

$$\rho U\ddot{a}=cAUa \quad 即 \quad \ddot{a}=\frac{c}{\rho}\mathscr{A}a.$$

这就是简正模式所满足的方程组(6).

现在我们来求解方程组(2).为此，以(5)式代入(1)右，并把(5)中的 $k$ 换成 $k'$，代入以后，对方程双方各乘以 $u_n^k$，并对 $n$ 求和，这时，利用(9)式，有

$$\sum_n \ddot{q}_n(t)u_n^k=\frac{c}{\rho}\sum_{k'}\sum_n a_{k'}(t)\bar{u}_n^k(u_{n+1}^{k'}-2u_n^{k'}+u_{n-1}^{k'}).$$

但是由(7)有 $u_{n\pm1}^{k'}=\mathrm{e}^{\pm ik'\alpha}u_n^{k'}$.利用(5)对 $t$ 求导二次，再代入上式左，利用(9)有

$$\ddot{a}_k(t)=\frac{c}{\rho}\sum_{k'}a_{k'}\sum_n \bar{u}_n^k u_n^{k'}(\mathrm{e}^{ik'\alpha}-2+\mathrm{e}^{-ik'\alpha}).$$

对上式右方再用一次(9)，上式右方成为

$$\frac{c}{\rho}(\mathrm{e}^{ik\alpha}-2+\mathrm{e}^{-ik\alpha})a_k(t)=-\omega_k^2a_k(t),$$

$$|\omega_k|=\sqrt{\frac{2c}{\rho}(1-\cos k\alpha)}=2\sqrt{\frac{c}{\rho}}\left|\sin\frac{k\alpha}{2}\right|.$$

所以简正坐标 $a_k(t)$ 适合一个完全解耦的方程组

$$\ddot{a}_k(t)+\omega_k^2a_k(t)=0,\quad \omega_k=2\sqrt{\frac{c}{\rho}}\left|\sin\frac{k\alpha}{2}\right|. \tag{11}$$

方程(11)即所谓谐振子(harmonic oscillator)方程，它的解就是本书第二章一开始所介绍的指数—三角函数

$$\begin{aligned}a_k(t)&=b_k\mathrm{e}^{-i\omega_k t}+\overline{b_{-k}}\mathrm{e}^{i\omega_k t}\\&=A_k\mathrm{e}^{-i(\omega_k t+\theta_k)},\end{aligned} \tag{12}$$

这里我们取系数为 $b_k$ 与 $\overline{b_{-k}}$ 目的在于保证 $\overline{a_k}(t)=a_{-k}(t)$，而这是保证最终 $q_n(t)$ 取实值之所必需.这样我们看到，杆的纵振动的离散模型表明：一个 $N$ 自由度的物理系统的运动实由一些谐振子构成.这些谐振子的频率 $\omega_k$ 全由这个物理系统决定：决定它的关系式 $\omega_k=2\sqrt{\frac{c}{\rho}}\cdot\left|\sin\frac{k\alpha}{2}\right|$ 称为色散关系，在研究这个物理系统的本性时起重要的作用.这些谐振子的振幅 $A_k$ 与初始位相则由初始条件决定.一根杆的纵振动可以是多种多样的，但总是由这些简正模式合成的.所以，分析一个物理系统的状况就应该首先求出这些简正模式，这些简正模式的频率就称为这个物理系统的谱(spectrum)，它刻画了这个系统的物理本质.上面讲到的寻求"隐藏着的周期性"就是这个意思.至于一个系统是否有如上类型的隐藏着的周期性，在我们的例子中全决定于一个算子(矩阵就是一个算子)可否对角化.如果可以，我们就看到，谱就是特征根.如果不

行，问题将极为复杂.

重要的是要提到，我们虽然只讲了一个机械运动的例子——杆的纵振动，但是这样的问题在研究种种自然现象时都会遇到，特别是在量子力学中. 由于不假设读者知道很多量子力学知识，我们只能限于经典物理学中的问题. 现在只提一下为什么我们就杆的离散模型讲这么多？一根连续的杆就是 $N\to\infty$时的极限状况，它是一个具有无穷自由度的物理系统. 当过渡到这个极限状况时，许多有趣的性质将会被掩盖，特别是色散关系. 但是当傅里叶着手研究热传导问题时，他确实是从极限状况开始的. 下面我们就仿照傅里叶的方法来考虑连续杆的纵振动. 我们从方程(1)开始，并令 $a\to 0$，由于(1)中 $c=K/a^2$ 而$\frac{1}{a^2}[q_{n+1}-2q_n+q_{n-1}]\to\frac{\partial^2 q}{\partial x^2}$，所以代替常微分方程组(1)，我们会得到一个偏微分方程：

$$\rho\frac{\partial^2 q}{\partial t^2}=K\frac{\partial^2 q}{\partial x^2},\quad 或写作\frac{1}{a^2}\frac{\partial^2 u}{\partial t^2}=\frac{\partial^2 u}{\partial x^2}.$$

这里 $a^2=K/\rho$，它不是上面讲到的 $a$，写成 $a^2$ 只不过是为了保证它取正值. 除了方程以外还应该有边值条件，如果研究两端固定的杆，就有 $q(0,t)=q(l,t)=0$；如果研究周期情况，则有 $q(0,t)=q(L,t)$. 此外，因为(1)是牛顿的动力学方程，求解时应知道初位移与初速，它的连续极限，也就成为相应的初始条件 $q(x,0)=\varphi(x)$，$q_t(x,0)=\psi(x)$. 总之，我们会有一个如下的初边值问题(我们把未知函数改写为 $u(x,t)$了)：

$$\begin{cases}\frac{1}{a^2}\frac{\partial^2 u}{\partial t^2}=\frac{\partial^2 u}{\partial x^2}, t\geqslant 0,\\ u(0,t)=u(L,t)=0, 0\leqslant x\leqslant L,\\ u(x,0)=\varphi(x),\ u_t(x,0)=\psi(x).\end{cases}\tag{13}$$

(13)中的偏微分方程通常称为弦振动方程. 这是因为，和常见的情况一样，数学物理中同样的方程可以刻画不同的物理系统，弦的横振动也适合同样的方程. 这个方程也有很长久的历史了. 达朗贝尔研究过它，丹尼尔•伯努利也研究过它，而且其解法与傅里叶的方法是一样的. 只不过傅里叶作了更系统的贡献. 这个方法如下：我们来看(5)式，这里 $q(t)$写成了一和式，其每一项都是 $t$ 的函数 $a_k(t)$与一个与 $t$ 无关，但与 $n$ 有关的 $u_n^k$ 相乘，因为 $n$ 代表离散模型中质点的位置，所以是一个空间变量，这样，在连续模型中我们试图求(13)的这样形状的解：

$$u(x,t)=X(x)T(t).$$

这里我们当然设 $X(x)\neq 0$，$T(t)\neq 0$，否则我们将会得到平凡不足道的解：$u\equiv 0$. 以 $u$ 的这个表达式代入(13)中的偏微分方程，即有

$$\frac{X''(x)}{X(x)}=\frac{1}{a^2}\frac{T''(t)}{T(t)},$$

所以它应该是一个常数 $-\lambda$. 于是对 $X(x)$和 $T(t)$我们分别得到以下的线性常微分方程：

$$X''(x)=-\lambda X(x),\ T''(t)=-\lambda a^2T(t).$$

暂将后一方程放在一边，先看第一个方程. 我们想求它的一个非零解 $X(x)\neq 0$，但希望这样作出的 $u(x,t)$能适合(13)中的边值条件. 因此要求

$$\begin{cases}X''+\lambda X=0,\\ X(0)=X(L)=0.\end{cases}\tag{14}$$

如果 $\lambda=-\mu^2<0$，则上之方程有解

$$X(x)=A\mathrm{e}^{\mu x}+B\mathrm{e}^{-\mu x}.$$

但因要求 $X(0)=X(L)=0$，就必须有 $A=B=0$，只能得到无意义的零解. 所以 $\lambda=-\mu^2<0$ 是不可能的. 若 $\lambda=0$，同样又有 $X(x)=A+Bx$，而由边值条件又得到零解. 因此为使上式有解，必须 $\lambda=\mu^2>0$，但是又非一切正的 $\lambda$ 都可以给出非零的解：当 $\lambda=\mu^2>0$ 时，$X(x)=A\cos\mu x+B\sin\mu x$，由 $X(0)=0$，即有 $A=0$，再由 $X(L)=0$，又有 $B\sin\mu L=0$. 但是 $B=0$ 又是没有用的. 因此 $\mu$ 必须取得使 $\sin\mu L=0$，从而 $\mu=k\pi/L$，$k=1,2,3,\cdots$，($k=-1,-2,-3,\cdots$给出同样的 $X(x)$). 总之，我们要求参数 $\lambda$ 之值使(14)有非零解. 这个问题称为斯图姆-刘维尔问题(Sturm-Liouville 简称 S-L 问题)，即求 $\lambda$ 之特定值，使(14)有非零解 $X(x)$. 所得的 $\lambda=(k\pi/L)^2$ 称为其固有值(eigenvalue)，相应的 $X(x)=B\sin\frac{k\pi}{L}x$ 称为相应的固有函数(eigenfunctions). 固有值的全体构成 S-L 问题之谱. 固有值与固有函数在许多文献中分别称为特征值与特征函数. 它与矩阵的特征根实质一样. 于是我们看到 S-L 问题(14)之谱即一组实的、分离的固有值之集合，所以我们说这个问题有离散的谱. 固有函数 $X(x)$ 中含有一个特定常数 $B$，我们将这样去取它，使 $X(x)$ 的 $L^2$ 范数为 1，在我们的情况下，因为

$$\int_0^L \sin^2\frac{k\pi x}{L}\mathrm{d}x=\frac{1}{2}\int_0^L\left(1-\cos\frac{2k\pi x}{L}\right)\mathrm{d}x=\frac{L}{2},$$

故常数 $B^2=\frac{2}{L}$，物理学家喜欢称它为归一化因子.

在解出了 $X(x)$ 后，再来看 $T(t)$. 由 $T''(t)+\lambda a^2T=0$ 以及 $\lambda=\mu^2=\left(\frac{k\pi}{L}\right)^2$ 即有

$$\begin{aligned}T(t)&=C\cos\frac{k\pi at}{L}+D\sin\frac{k\pi at}{L}\\&=M_k\cos\left(\frac{k\pi at}{L}+\theta_k\right).\end{aligned}$$

这又是一个谐振子，其频率是$\frac{k\pi a}{L}$(相当于离散模型中的 $\omega$，但是现在的 $a$ 相当(1)中的$a\sqrt{\frac{c}{\rho}}$)在讨论离散模型时，我们的频率 $\omega_k$ 与两个质点之间的距离 $a$ 之间有色散关系 $\omega_k=2\sqrt{\frac{c}{\rho}}\cdot\left|\sin\frac{ka}{2}\right|$. 如果 $a\to0$ 因为 $\sin\frac{ka}{2}\sim\frac{ka}{2}$，所以 $\omega_k\sim a\sqrt{\frac{c}{\rho}}k$，即在极限情况下 $\omega_k$ 与 $k$ 成正比，色散关系特别简单，称为非色散波. 在连续情况下，正是 $\omega_k\sim k$ 这个简单式子把一般的较为复杂的色散关系掩盖起来了.

关于振动与波的讨论我们暂时只能停在这里了，下面再回到傅里叶级数的讨论.

初边值问题(13)中方程与边值条件俱已满足，余下的只有初值条件. 为满足它们，需要把所得的 $X(x)T(t)$叠加起来，得

$$u(x,t)=\sum_{k=0}^{\infty}\left(C_k\cos\frac{k\pi at}{L}+D_k\sin\frac{k\pi at}{L}\right)\sin\frac{k\pi x}{L},$$

这里 $C_k$ 与 $D_k$ 都是未定数. 我们需要这样确定它们，使得

$$\varphi(x)=\sum C_k\sin\frac{k\pi x}{L},\psi(x)=\sum_{k=0}^{\infty}\frac{k\pi a}{L}D_k\sin\frac{k\pi x}{L}. \tag{15}$$

从初等的微积分教本中我们已熟知，它们正是 $\varphi(x)$ 和 $\psi(x)$ 的傅里叶展开式，而且

$$C_k=\frac{2}{L}\int_0^L\varphi(x)\sin\frac{k\pi x}{L}\mathrm{d}x,$$

$$D_k=\frac{2}{k\pi a}\int_0^L\psi(x)\sin\frac{k\pi x}{L}\mathrm{d}x.$$

由此以下我们转入纯粹数学的讨论.

**2. 傅里叶级数及其收敛性问题**　第四章一开始我们就指出了，傅里叶由此提出了任意函数都可以展开为傅里叶级数的问题，这对于从 18 世纪以来就存在的何为函数的争论正如火上加油. 其实从 18 世纪以后，许多数学大师都参加进来，互相问难以至"诋毁"，终于有了狄利克雷的函数定义. 黎曼积分的出现与此问题直接相关，勒贝格积分理论也是这样. 所以许多人说，傅里叶分析是现代分析数学的起点. 但是它不仅改变了整个分析数学的面貌，而且与群表示理论，代数数论、概率论等等都密切相关，出现了调和分析这个广大的分支学科. 集合论的研究也始自这个问题，而集合论的出现改变了全部数学的基础这是没有疑问的了.

这个问题的中心，首先是函数 $f(x)$ 在什么条件下可以展开为其傅里叶级数？因此，本节前半部分主要是讨论傅里叶级数如何由谱的问题产生，下面则要讨论其中心问题——收敛性问题. 以下为简单起见，我们在区间 $[-\pi,\pi]$ 上讨论 $f(x)$ 之傅里叶级数，而且采用复记号：

$$f(x)\sim\sum_{k=-\infty}^{\infty}C_k\mathrm{e}^{\mathrm{i}kx},$$

$$C_k=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x. \tag{16}$$

它与常见的实的记法

$$f(x)\sim\frac{a_0}{2}+\sum_{k=1}^{\infty}(a_k\cos kx+b_k\sin kx);$$

$$a_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos kx\mathrm{d}x,\quad b_k=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin kx\mathrm{d}x \tag{17}$$

不同：在 (16) 中 $k$ 从 $-\infty$ 变到 $+\infty$，而在 (17) 中 $k$ 则从 0 变到 $+\infty$，而且

$$a_k=C_k+C_{-k},\quad b_k=\mathrm{i}(C_k-C_{-k}). \tag{18}$$

所以积分号前的因子在 (16) 中是 $\frac{1}{2\pi}$，在 (17) 中则是 $\frac{1}{\pi}$.

(16) 或 (17) 都是适用于 $[-\pi,\pi]$ 上的. 但其右方则是 $\mathbf{R}$ 上的以 $2\pi$ 为周期的函数，所以在 $[-\pi,\pi]$ 外也是有定义的，而 $f(x)$ 则只在 $[-\pi,\pi]$ 上有定义. 因此我们说 (16) 或 (17) 代表 $f(x)$ 在 $\mathbf{R}$ 上以 $2\pi$ 为周期的周期延拓. 准确些说则是把 $f(x)$ 在 $[-\pi,\pi]$ 中的一段周期地重现到 $\mathbf{R}$ 上. (16) 或 (17) 右方以任一长为 $2\pi$ 的区间为基本周期，不论这个基本周期是 $[0,2\pi]$ 或 $[-\pi,\pi]$，只是平移一下 $\mathrm{e}^{\mathrm{i}kx}$ 之图像而已. 但若 $f(x)$ 原来就定义在 $\mathbf{R}$ 上，取其 $[-\pi,\pi]$ 上的一段或取其 $[0,2\pi]$ 上的一段作周期延拓，其结果就可以完全不同. 所以作 $f(x)$ 之傅里叶展开 (16) 或 (17) 时必须先说清是取 $f(x)$ 之"哪一段"来作展开. 我们特别要注意，即令 $f(x)$ 在 $\mathbf{R}$ 上连续，取其 $[-\pi,\pi]$ 上之一段作周期延拓后，可能得到不连续的函数. 例如若记周期延拓后所得函数为

$f^*(x)$，则 $f^*(\pi-0)=f(\pi)$，$f^*(\pi+0)=f(-\pi)$，所以在 $x=\pi$ 处会出现一个第一类间断点．对 $f(x)$之各阶导数也是这样．这一点下面应该特别注意．

既然得出了 $f(x)$之傅里叶展开式(16)或(17)，基本的问题自然就是收敛性问题．但我们不妨从更宽的角度来看待它，而问这些级数与 $f(x)$关系如何？例如它们是否收敛，若收敛是否其和一定为 $f(x)$等等．要注意，收敛性问题只是问题的一部分，所以我们在(16)或(17)中都未用"="号而用"~"．首先是，我们打算在什么框架下讨论这些问题．在 19 世纪中叶狄利克雷和黎曼的时代，自然是在黎曼可积函数类中讨论它．如果没有适当的可积性理论，(16)与(17)中的傅里叶系数公式就无法定义，而且正如第四章一开始就指出的，黎曼积分理论正是这样出来的．到 19 世纪末，有了勒贝格积分理论(它至少部分地也是由傅里叶级数理论引起的)，人们自然会在 $L^1$ 框架下讨论它．$L^2$ 空间中的傅里叶级数理论我们已在第四章§3 中讲了一个头．它后来引导到泛函分析、数学物理等很广阔的天地中去了，我们也就不再讨论它了．$L^1$ 与 $L^2$ 中的傅里叶级数理论都极为细致，由于我们分析数学方面准备不足，只好割爱．在本章中，我们较多地从广义函数角度来讨论它，并适当涉及 $L^1$ 与 $L^2$ 理论．

在进一步讨论(16)或(17)的收敛性前，我们还要作一个约定，因为(16)可以直接由(17)经过关系式(18)导出，所以若对(17)取部分和 $k=0,1,\cdots,N$，则(16)中也只出现 $\sum_{k=-N}^{N}$．因此以下凡讲到(16)的部分和时恒指

$$S_N(x)=\sum_{k=-N}^{N}C_k\mathrm{e}^{\mathrm{i}kx}. \tag{19}$$

为了讨论(16)或(17)的收敛性，一个最直接了当的办法是估计 $C_k$ 的大小．这里我们有一个基本的，然而很简单的

**定理 2**　设 $f(x)$由其定义域$[-\pi,\pi]$经过以 $2\pi$ 为周期的延拓后成为一个 $C^m$ $(m>1)$类函数，则其傅里叶系数 $C_k$(或 $a_k$，$b_k$)适合估计式

$$|C_k|\leqslant M/k^m\ (|a_k|,|b_k|\leqslant M/k^m), \tag{20}$$

这里 $M$ 与 $k$ 无关．

**证**　因为 $f(x)$在周期延拓后为 $C^m$ 函数，所以 $f(\pi-0)=f(\pi)=f(-\pi)=f(-\pi+0)$．对 $f$ 的不高于 $m$ 阶的导数，也有这样的公式．由分部积分法

$$\begin{aligned}C_k&=-\frac{1}{2k\pi\mathrm{i}}f(x)\mathrm{e}^{-\mathrm{i}kx}\Big|_{-\pi}^{\pi}+\frac{1}{2k\pi\mathrm{i}}\int_{-\pi}^{\pi}f'(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x\\&=\frac{1}{2k\pi\mathrm{i}}\int_{-\pi}^{\pi}f'(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x\\&=\cdots=\frac{1}{2\pi(k\mathrm{i})^m}\int_{-\pi}^{\pi}f^{(m)}(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x.\end{aligned}$$

上面我们多次应用了分部积分．由此直接得出(20)．

这个定理看来简单，但若 $m\geqslant 2$，由它可以得证(16)或(17)绝对而一致地收敛(有些文献上称为正规收敛)，而这个结果对于很一般的 S-L 问题也是适用的，在那里我们没有办法应用指数函数和三角函数那么丰富的性质．但是即令是如此简单的定理，也还有许多值得挖掘的地方．例如，如果 $f(x)$ 在 $x=\pi$ 处有第一类间断点，则

$-\frac{(-1)^k}{2k\pi\mathrm{i}}[f(-\pi+0)-f(\pi-0)]=\frac{(-1)^{k+1}}{2k\pi\mathrm{i}}[f(\pi+0)-f(\pi-0)]$(注意 $f(x)$ 以 $2\pi$ 为周期)，从而 $C_k=O\left(\frac{1}{k}\right)$，反过来，若 $C_k=O\left(\frac{1}{k}\right)$，我们可以猜到，函数会有第一类间断点. 本节下面的例子说明确是如此. 因此产生了一个根本思想：$f(x)$ 的光滑性与其傅里叶系数衰减速度直接相关. 这个重要思想将在本章最后一节中详细讨论. 现在我们先要再讲一下 $L^p$ 函数的一个性质.

我们在上一章中已经看到，任一个 $L^p$ 函数 $(1\leqslant p<+\infty)$ 都可以用一个连续函数去按 $L^p$ 系数逼近. 但我们没有详细证明，这不是因为这一性质不重要，恰好相反，它是极为重要的，因此我们将在下一章在子空间的稠密性的背景下去讨论它. 我们这里只指出，$f(x)\in L^p$ 不但可以用连续函数去逼近，而且可以通过例如磨光算子，用 $C_0^\infty$ 函数，即 $D$ 中的函数去作 $L^p$ 逼近. 不但如此，我们还可以用阶梯函数去逼近. 所谓区间 $I$ 上的阶梯函数，即把 $I$ 分割成有限多个小区间：

$$I: a=a_0<a_1<\cdots<a_N=b \quad (I \text{ 即}[a,b]),$$

而阶梯函数即在每个小区间上各取常数值的函数：

$$\varphi(x)=A_i \quad a_{i-1}\leqslant x\leqslant a_i \quad i=1,2,\cdots,N.$$

此外还可以证明 $f(x)\in L^p$ 有所谓平均连续性，即对任一 $\varepsilon>0$ 必可找到一个 $\delta$，使当 $|h|<\delta$ 时

$$\|f(x+h)-f(x)\|=\left(\int_I|f(x+h)-f(x)|^p\mathrm{d}x\right)^{1/p}<\varepsilon.$$

这些性质不仅对有界的 $I$ 成立，对"$\sigma$ 可测"的 $I$ 例如 $I=\mathbf{R}$ 也是成立的. 下面我们将自由地应用这些结果.

首先给出一个重要的定理.

**定理 3**(黎曼－勒贝格引理)　若 $f(x)\in L^1(\mathbf{R})$ 则 $\int_{-\infty}^{\infty}f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x\to 0 \quad (k\to\infty)$.

**证**　因为 $f(x)\in L^1(\mathbf{R})$，故对任意 $\varepsilon>0$ 必可找到正数 $M$，使

$$\int_M^{+\infty}|f(x)|\mathrm{d}x,\int_{-\infty}^{-M}|f(x)|\mathrm{d}x<\frac{\varepsilon}{3}.$$

所以

$$\int_{-\infty}^{\infty}f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x=\left[\int_{-\infty}^{-M}+\int_M^{+\infty}\right]f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x+\int_{-M}^{M}f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x,$$

其中前两项绝对值均小于 $\varepsilon/3$，现在看第三项. 由前述 $L^p$ 函数的性质，可以找到一个阶梯函数 $w(x)$，使

$$\int_{-M}^{M}|f(x)-w(x)|\mathrm{d}x<\frac{\varepsilon}{6},$$

而如定理 2 前所述 $\int_{-M}^{M}w(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x$ 由有限项 $A_j\int_{a_{j-1}}^{a_j}\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x=\frac{\mathrm{i}}{k}A_j[\mathrm{e}^{-\mathrm{i}ka_j}-\mathrm{e}^{-\mathrm{i}ka_{j-1}}]$相加而得，当 $k$ 充分大时它们可以变得任意小. 所以只要 $k$ 充分大，必有

$$\left|\int_{-M}^{M}f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x\right|\leqslant\int_{-M}^{M}|f(x)-w(x)||\mathrm{e}^{-\mathrm{i}kx}|\mathrm{d}x$$
$$+\sum_{j=1}^{N}\left|A_j\int_{a_{j-1}}^{a_j}\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x\right|<\frac{\varepsilon}{3}.$$

因此当 $k$ 充分大时

$$\left|\int_{-\infty}^{\infty} f(x)\mathrm{e}^{\mathrm{i}kx}\mathrm{d}x\right| \leqslant \left[\int_{-\infty}^{-M} + \int_{M}^{+\infty}\right]|f(x)\mathrm{e}^{-\mathrm{i}kx}|\mathrm{d}x + \left|\int_{-M}^{M} f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x\right| < \varepsilon.$$

特别是，若 $f(x)\in L^1([-\pi,\pi])$，则 $f(x)$ 之傅里叶系数 $C_k$ 当 $k\to\infty$ 时，必趋于 0.

由于这个定理的重要性，我们不妨分析一下其要点何在. 实际上，重要的是右方第三个积分的估计：我们用 $w(x)$ 去逼近 $f(x)$，而 $w(x)$ 至少在一个小区间 $[a_{j-1},a_j]$ 上是不变的. 而另一个因子 $\mathrm{e}^{\mathrm{i}kx}$，因为其周期为 $\dfrac{2\pi}{k}$，这个小区间中包含了它的许多周期. 每一个周期中 $\mathrm{e}^{\mathrm{i}kx}=\cos kx+\mathrm{i}\sin kx$ 之实虚部都是正号地变一段，又完全同样地反号再变一段，因此乘上 $w(x)=A_j$ 后，平均地完全抵消了. $f(x)$ 虽然不如 $w(x)$ 一样在某区间中不变，但既可用 $w(x)$ 逼近，也就可以说是缓变的. 所以 $\int_{-M}^{M} f(x)\mathrm{e}^{\mathrm{i}kx}\mathrm{d}x$ 中既有缓变因子 $f(x)$，又有急变因子 $\mathrm{e}^{-\mathrm{i}kx}$；二者相乘又相加，就好像是使 $f(x)$ 之值加“权”正负相消了(这里用语不太准确，因为“权”一般总是正的). 这实际上是傅里叶级数的一个“诀窍”：例如在 $\sum C_k\mathrm{e}^{\mathrm{i}kx}$ 中 $\mathrm{e}^{\mathrm{i}kx}$ 当 $k$ 充分大时是急变的，即高频成分. 我们证明定理 2 时只利用了关于 $|C_k|$ 的估计，而对 $\mathrm{e}^{\mathrm{i}kx}$ 之振荡特性用到不多，所以得到的结果必然是比较粗糙的.

但是即令如此，我们通过分部积分以及 $\int \mathrm{e}^{\mathrm{i}kx}\mathrm{d}x=\dfrac{1}{\mathrm{i}k}\mathrm{e}^{\mathrm{i}kx}+C$ 仍然利用了 $\mathrm{e}^{\mathrm{i}kx}$ 之振荡特性. 话已至此，索性多讲几句在今后可能会对读者有用的话.

**3. 阿贝尔变换和第二积分中值定理**　我们处理的级数是形如 $\sum a_kb_k$ 的，分部积分讨论的是 $f(x)\cdot\varphi(x)$ 的积分，如果暂时不谈极限，则也是 $\sum\limits_{k=1}^{N} a_kb_k$ 类型的和式. 关于这类和式有一个著名的阿贝尔变换：如果我们记 $B_m=\sum\limits_{k=1}^{m} b_k$，则 $b_k=B_k-B_{k-1}$(规定 $B_0=0$)，因此

$$\begin{aligned} S &= \sum_{k=1}^{N} a_kb_k = \sum_{k=1}^{N} a_k(B_k-B_{k-1}) \\ &= \sum_{k=1}^{N-1}(a_k-a_{k+1})B_k + a_NB_N. \end{aligned} \tag{21}$$

现在，把这个变换用于无穷级数 $\sum\limits_{k=1}^{\infty} a_kb_k$. 为了证明它的收敛，只需应用柯西准则，即对 $\varepsilon>0$ 找一个 $M$，使得对任意的 $N$

$$\left|\sum_{i=M+1}^{M+N} a_ib_i\right| = \left|\sum_{k=1}^{N} a_{M+k}b_{M+k}\right| < \varepsilon.$$

对此式则可再应用上述阿贝尔变换而有以下两个判别法：

**定理 4**　(i) **阿贝尔判别法**　若 $\sum\limits_{h=1}^{\infty} b_k$ 收敛而 $\{a_k\}$ 是单调有界序列，则上述级数收敛.

(ii) **狄利克雷判别法**　若 $\sum\limits_{k=1}^{\infty} b_k$ 的部分和序列有界，而 $\{a_k\}$ 单调趋于 0，则上述级数收敛.

**证**　我们把(21)应用于 $\sum\limits_{i=M+1}^{M+N} a_ib_i$，于是令 $B_k=\sum\limits_{i=1}^{k} b_{M+i}$. 先看

对于(i)，这时只要 $M$ 充分大则一切 $B_k$，$B_N$ 都适合 $|B_k|<\varepsilon$，又因 $\{a_k\}$ 有界，所以 $|a_k|\leqslant A$，从而 $|a_k-a_{k+1}|\leqslant|a_k|+|a_{k-1}|\leqslant 2A$，$|a_N|\leqslant A$，应用柯西准则，

$$|S|=\left|\sum_{k=1}^{M} a_{M+k}b_{M+k}\right|\leqslant 2A\varepsilon+A\varepsilon=3A\varepsilon.$$

从而原级数收敛.

在(ii)的情况，一切 $|B_k|\leqslant B$，而 $|a_{M+k}|<\varepsilon$，应用柯西准则，又有

$$|S|\leqslant 2B\varepsilon+B\varepsilon=3B\varepsilon.$$

定理证毕.

阿贝尔变换在积分中有什么类比吗？首先就是分部积分法. 现在我们不要太严格，这样可以自由地在积分和和定积分之间随意变换. 考虑 $\int_I f(x)\varphi(x)\mathrm{d}x$，通过把 $I$ 分割成许多小区间 $[x_0,x_1]\cup[x_1,x_2]\cup\cdots\cup[x_{N-1},x_N]$把它变成积分和 $\sum_{k=1}^{N} f(\xi_k)\varphi(\xi_k)\Delta_k x$，令 $a_k=f(\xi_k)$，$b_k=\varphi(\xi_k)\Delta_k x$，则 $B_m=b_1+\cdots+b_m=\sum_{k=1}^{m}\varphi(\xi_k)\Delta_k x$，我们又把它写成 $B(x_m)=\int_{x_0}^{x_m}\varphi(x)\mathrm{d}x$（注意 $B(x_0)=0$），于是阿贝尔变换成了

$$\begin{aligned}\int_I f(x)B'(x)\mathrm{d}x &= \sum\ [a(\xi_k)-a(\xi_{k+1})]\int_{x_0}^{x_k}\varphi(x)\mathrm{d}x\\ &\quad + a(x_N)\int_{x_0}^{x_N}\varphi(x)\mathrm{d}x.\end{aligned}\tag{22}$$

但

$$a(\xi_{k-1})-a(\xi_{k+1})=-\int_{\xi_{k-1}}^{\xi_{k+1}} f'(x)\mathrm{d}x,$$

忽略一些“可以忽略的”误差，还注意到 $B(x_0)=0$，并把 $x_N$ 写成 $x$，阿贝尔变换式(22)立即成了分部积分公式

$$\int_I f(x)\varphi(x)\mathrm{d}x=f(x)B(x)\Big|_{x_0}^{x}-\int_{x_0}^{x} f'(x)B(x)\mathrm{d}x.$$

我们在第四章中就说过积分(黎曼积分)其实就是组合学公式“加上一点极限”. 达布关于牛顿-莱布尼茨公式的证明说明了它，分部积分法又一次说明了它. 可是，把阿贝尔变换用于积分学对我们当前讨论的傅里叶级数最重要的贡献在于它给出了第二积分中值定理.

**定理 5**(**第二积分中值定理**)　设 $f(x)$ 与 $\varphi(x)$ 均定义在 $[a,b]$ 上，$f(x)$ 单调而 $\varphi(x)$ 连续，则必存在 $\xi\in[a,b]$，使

$$\int_a^b \varphi(x)f(x)\mathrm{d}x=f(a+0)\int_a^{\xi}\varphi(x)\mathrm{d}x+f(b-0)\int_{\xi}^{b}\varphi(x)\mathrm{d}x.\tag{23}$$

**证**　不失一般性可以设 $f(x)$ 单调下降(否则考虑 $-f(x)$). 单调函数最多有可数多个第一类间断点而且必可积. 于是 $\varphi(x)f(x)$ 可积. 对单调函数 $f(x)$，$f(a+0)$，$f(b-0)$ 均存在，我们又可设 $f(b-0)=0$，否则可以用 $F(x)=f(x)-f(b-0)$ 代替 $f(x)$. 如果我们证明了

$$\int_a^b \varphi(x)F(x)\mathrm{d}x = F(a+0)\int_a^\xi \varphi(x)\mathrm{d}x + F(b-0)\int_\xi^b \varphi(x)\mathrm{d}x$$
$$= F(a+0)\int_a^\xi \varphi(x)\mathrm{d}x, \tag{24}$$

将 $F(x)=f(x)-f(b-0)$代入上式即可得到(23).

现在作$[a,b]$的一个分划，$a=x_0<x_1<\cdots<x_N=b$，并令 $G(x)=\int_a^x \varphi(x)\mathrm{d}x$ 为$\varphi(x)$的一个原函数，则 $G(x)$ 是 $[a,b]$ 上的连续函数，从而有最小值 $m$ 和最大值 $M$. 我们要作 $\int_a^b F(x)\varphi(x)\mathrm{d}x$ 的一个积分和 $\sum_{i=1}^{N} F(\xi_i)\varphi(\xi_i)\Delta_i x$. $\xi_i$ 的取法如下：对下式应用积分中值定理，

$$G(x_i)-G(x_{i-1})=\int_{x_{i-1}}^{x_i} \varphi(x)\mathrm{d}x=\varphi(\xi_i)\Delta_i x,$$

并且利用这样定出的 $\xi_i$ 作上述积分和，于是对 $\sum_{i=1}^{N} F(\xi_i)\varphi(\xi_i)\Delta_i x$ 应用阿贝尔变换，而以 $F(\xi_i)$作为 $a_i$，$\varphi(\xi_i)\Delta_i x$ 作为 $b_i$，于是 $B_i=b_1+\cdots+b_i=\sum_{j=1}^{i}\int_{x_{j-1}}^{x_j} \varphi(\xi_i)\mathrm{d}x=\int_a^{x_i}\varphi(x)\mathrm{d}x=G(x_i)$. 应用阿贝尔变换有

$$\sum_{i=1}^{N} F(\xi_i)\varphi(\xi_i)\Delta_i x=\sum_{k=1}^{N-1}[F(\xi_k)-F(\xi_{k+1})]B_k+F(\xi_N)B_N.$$

因为 $F(x)$是单调下降而且非负的，$B_k$ 与$B_N$ 均为$G(x)$之值，从而 $m\leqslant B_k, B_N\leqslant M$，代 入上式有

$$m\left[\sum_{k=1}^{N-1}(F(\xi_k)-F(\xi_{k+1}))+F(\xi_N)\right]\leqslant \sum_{i=1}^{N} F(\xi_i)\varphi(\xi_i)\Delta_i x$$
$$\leqslant M\left[\sum_{k=1}^{N-1}(F(\xi_k)-F(\xi_{k+1}))+F(\xi_N)\right],$$

亦即

$$mF(\xi_1)\leqslant \sum_{i=1}^{N} F(\xi_i)\varphi(\xi_i)\Delta_i x\leqslant MF(\xi_1).$$

取极限后即得

$$mF(a+0)\leqslant \int_a^b F(x)\varphi(x)\mathrm{d}x\leqslant MF(a+0).$$

由于 $m, M$ 分别是$\varphi(x)$在$[a,b]$上的最小与最大值，由连续函数的中间值定理，一定存在 $\xi\in[a,b]$使 $\varphi(\xi)$之值适合

$$\int_a^b F(x)\varphi(x)\mathrm{d}x=\varphi(\xi)F(a+0).$$

此即(24)式，从而定理成立.

与关于无穷级数的收敛性相应，关于积分$\int_a^{+\infty} f(x)\varphi(x)\mathrm{d}x$，我们也有相应定理.

**定理 4′ (i)阿贝尔判别法**　若积分$\int_a^{+\infty} f(x)\mathrm{d}x$ 收敛，但不一定绝对收敛，而 $\varphi(x)$单调有界，

则 $\int_a^{+\infty} f(x)\varphi(x)\mathrm{d}x$ 收敛.

(ii)**狄利克雷判别法** 若 $f(x)$ 在 $[a,+\infty]$ 的任一子区间 $[a,A]$ 上可积，而且 $\int_a^A f(x)\mathrm{d}x$ 对 $A$ 有界；$\varphi(x)$ 当 $x\to+\infty$ 时单调趋于 0，则 $\int_a^{+\infty} f(x)\varphi(x)\mathrm{d}x$ 收敛.

这个定理的证明我们略去. 但要指出，对其它类型的反常积分，也有同样的结果成立.

**4. 从广义函数观点看傅里叶级数的收敛性** 下面我们再回到对傅里叶级数的讨论. 我们来看级数(16)或(17)是否在 $x\in[-\pi,\pi]$ 处收敛于 $f(x)$. 这里假设已对 $f(x)$ 作了周期延拓. 把(16)中 $C_k$ 的公式(积分变量换成 $u$)代入(16)中的级数，则其部分和是

$$\begin{aligned}S_N(x)&=\frac{1}{2\pi}\int_{-\pi}^{\pi} f(u)\sum_{k=-N}^{N}\mathrm{e}^{-\mathrm{i}ku}\mathrm{e}^{\mathrm{i}kx}\mathrm{d}u\\&=\frac{1}{\pi}\int_{-\pi}^{\pi} f(u)\left[\sum_{k=1}^{N}\cos k(x-u)+\frac{1}{2}\right]\mathrm{d}u,\end{aligned}$$

但是

$$\begin{aligned}\sum_{k=1}^{N}\cos k(x-u)+\frac{1}{2}&=\left[\sum_{k=1}^{N}\frac{\cos k(x-u)\sin\frac{1}{2}(x-u)}{\sin\frac{1}{2}(x-u)}+\frac{1}{2}\right]\\&=\sum_{k=1}^{N}\frac{\sin\left[\left(k+\frac{1}{2}\right)(x-u)\right]-\sin\left[\left(k-\frac{1}{2}\right)(x-u)\right]}{2\sin\frac{1}{2}(x-u)}+\frac{1}{2}\\&=\frac{\sin\left[\left(N+\frac{1}{2}\right)(x-u)\right]}{2\sin\frac{1}{2}(x-u)}.\end{aligned}$$

令 $u-x=t$ 代入上式，有

$$S_N(x)=\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}f(x+t)\mathrm{d}t.$$

为了证明 $\lim\limits_{N\to\infty}S_N(x)=f(x)$，只需证明当 $N\to\infty$ 时

$$S_N(x)-f(x)=\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}[f(x+t)-f(x)]\mathrm{d}t \tag{25}$$

趋向 0 即可，这里我们需要倒用上面的求和程序，而有

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\mathrm{d}t=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k=-N}^{N}\mathrm{e}^{\mathrm{i}kt}\mathrm{d}t=1.$$

迄今为止，我们对 $f(x)$ 之光滑性未加限制. 为简单计，暂时固定 $x$，并令 $f(t)\in C^{\infty}$ 且支集在 $x$

附近，则有

$$\frac{1}{2\sin\frac{t}{2}}[f(x+t)-f(x)]\in L^1(-\pi,\pi).$$

因此利用黎曼－勒贝格引理即知 $\lim\limits_{N\to\infty}S_N(x)=f(x)$. 这样做似乎没有得到好处：我们的结果还未超过定理 2，而且还没有考虑到当 $x$ 接近 $[-\pi,\pi]$ 的端点时因 $f(x)$ 周期延拓而可能产生的新的第一类间断点. 但是它告诉我们，$\mathbf{R}$ 上的以 $2\pi$ 为周期的 $C^\infty$ 函数 $D_N(t)=\dfrac{1}{\pi}\dfrac{\sin\left(N+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}$ 构成一个 $\delta$ 序列，$D_N(t)$ 称为狄利克雷核，于是我们有

**定理 6**　$\{D_N(t)\}$ 是一个 $\delta$ 序列，亦即

$$D_N(t)\to\delta(t)\quad\text{in}\quad D'(\mathbf{R}).\tag{26}$$

这个定理给了我们一个新的视角：注意到 $D_N(t)$ 是由 $\sum\limits_{k=-N}^{N}\mathrm{e}^{\mathrm{i}k(x-u)}$ 用几何数列求和得出来的（我们只是为了不在复域中迂回，才把它直接化为余弦之和并用了积化和差分式，其实使用几何数列求和结果是一回事，读者自己不妨一试），所以可以写成

$$\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\mathrm{e}^{\mathrm{i}kt}=\delta(t)\quad\text{in}\quad D'(\mathbf{R}).$$

如果再注意到 $\langle\delta(t),\mathrm{e}^{\mathrm{i}kt}\rangle=1$，则上式左方又可以说成是 $\delta(t)$ 的傅里叶级数，定理 6 就可以改述为 $\delta(t)$ 之傅里叶级数在 $D'(\mathbf{R})$ 中收敛于 $\delta(t)$. 这样说法听起来固然颇有新意，但追究到底，因为 $D'$ 广义函数是 $D(\mathbf{R})$ 上的连续线性泛函，所以我们还只是证明了 $D(\mathbf{R})$ 函数 $f(x)$ 之傅里叶级数必收敛于 $f(x)$——还远远没有超出定理 2！

但是 $\delta(t)$ 不仅是 $D(\mathbf{R})$ 上的连续线性泛函，在第二章 §4 中就指出了 $\delta(t)$ 甚至是连续函数空间上的线性泛函：$\langle\delta,f\rangle=f(0)$ 对连续函数即已成立，上面说的 $\langle\delta(t),\mathrm{e}^{\mathrm{i}kt}\rangle=1$ 就是用的这个结果. 于是傅里叶级数收敛问题就变成：$D_N(t)$ 在什么样的空间中可以看成 $\delta$ 序列？为了回答这个问题，我们还是认为 $f(x)$ 定义在 $[-\pi,\pi]$ 上，并经周期延拓到 $\mathbf{R}^1$ 上，于是在 $x=\pm\pi$ 处会有多余的间断点. 现在来考虑

$$\begin{aligned}
&S_N(x)-\frac{1}{2}[f(x+0)+f(x-0)]\\
&\quad=\frac{1}{\pi}\int_0^\pi\frac{\sin\left(N+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\left[f(x+t)-\frac{1}{2}f(x+0)-\frac{1}{2}f(x-0)\right]\mathrm{d}t\\
&\quad\quad+\frac{1}{\pi}\int_{-\pi}^0\frac{\sin\left(N+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\left[f(x+t)-\frac{1}{2}f(x+0)-\frac{1}{2}f(x-0)\right]\mathrm{d}t\\
&\quad=\frac{1}{\pi}\int_0^\pi\frac{\sin\left(N+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\{[f(x+t)-f(x+0)]+[f(x-t)-f(x-0)]\}\mathrm{d}t.
\end{aligned}$$

今记

$$\Psi(x,t)=\frac{1}{2\pi\sin\frac{t}{2}}([f(x+t)-f(x+0)]+[f(x-t)-f(x-0)]).$$

固定 $x$ 并视 $\Psi$ 为 $t$ 的函数，于是若 $\Psi(x,t)$ 为 $L^1$ 函数即可利用黎曼－勒贝格引理证明级数(16)收敛，而且其和是$\frac{1}{2}[f(x+0)+f(x-0)]$，即左、右极限的均值. 注意到 $\sin\frac{t}{2}\sim\frac{t}{2}$，所以只需$\frac{f(x\pm t)-f(x\pm 0)}{t}\in L^1$ 即可. 但是有许多情况可以保证这一点，例如设 $f(t)$ 在 $t=x$ 处适合赫尔德条件

$$|f(x+h)-f(x)|\leqslant C|h|^{\alpha},0<\alpha\leqslant 1$$

($\alpha=1$ 时称此条件为利普希茨条件)即可.

从今天的角度来看，我们自然知道，真正的问题在于在一个尽可能宽的函数空间中寻找一个 $\delta$ 序列，就是找一个含参数的核. 这种核为数众多，我们下面还可以看一个很类似的核

$$W_N(t)=\frac{1}{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}. \tag{27}$$

它与狄利克雷核相近，因为

$$\int_0^{\pi}\frac{\sin\left(N+\frac{1}{2}t\right)}{2\sin\frac{t}{2}}\varphi(t)\mathrm{d}t=\int_0^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}\Phi(t)\varphi(t)\mathrm{d}t,$$

其中 $\Phi(t)=\frac{t/2}{\sin\frac{t}{2}}$，它在$[-\pi,\pi]$中是很光滑的. 我们用参数 $m$ 代替$N+\frac{1}{2}$，下面证明$\frac{1}{\pi}\frac{\sin mt}{t}$当 $m\to\infty$ 在满足狄利克雷条件的函数类中是 $\delta$ 序列，所谓狄利克雷条件即

$(D)$：若 $f(x)$在$[-\pi,\pi]$上最多只有有限多个第一类间断点，而且可将$[-\pi,\pi]$分为有限多个区间，使 $f(x)$在每个区间上均为单调，则称 $f(x)$适合条件$(D)$——狄利克雷条件.

**定理 7**　核 $W_N(t)$在上述函数类中是 $\delta$ 序列，即是说，若 $f(x)$适合条件$(D)$，则有

$$\lim_{N\to\infty}\frac{1}{\pi}\int_{-\pi}^{\pi}W_N(t)f(t)\mathrm{d}t=\frac{1}{2}[f(0+)+f(0-)]. \tag{28}$$

在证明此式前，我们先要提到一个著名的积分公式

$$\int_0^{\infty}\frac{\sin t}{t}\mathrm{d}t=\frac{\pi}{2}.$$

值得注意的是这个积分收敛但是并不绝对收敛. 因为比较详尽的微积分教材都有这个例子，我们就不来重复证明了. 现在我们回到定理 7 的证明.

我们考虑

$$\begin{aligned}&\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}\left[f(x+t)-\frac{1}{2}f(x+0)-\frac{1}{2}f(x-0)\right]\mathrm{d}t\\ =&\frac{1}{\pi}\int_0^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}[f(x+t)-f(x+0)]\mathrm{d}t\end{aligned}$$

$$+\frac{1}{\pi}\int_{-\pi}^{0}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}[f(x+t)-f(x-0)]\mathrm{d}t.$$

这里我们已利用了$\frac{\sin\left(N+\frac{1}{2}\right)t}{t}$是 $t$ 的偶函数，从而 $\int_{-\pi}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}\frac{1}{2}f(x+0)\mathrm{d}t=\int_{0}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}f(x+0)\mathrm{d}t$. 在后一积分中作变量变换 $t\to -t$，则上式可化为

$$\frac{1}{\pi}\int_{0}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}[f(x+t)+f(x-t)-f(x+0)-f(x-0)]\mathrm{d}t. \tag{29}$$

由条件(D)知，只要取 $\delta$ 充分小，$f(x)$在 $x$ 之左侧$[x-\delta,x]$和右侧$[x,x-\delta]$中均为单调的，又由于$\frac{\sin\left(N+\frac{1}{2}\right)t}{t}$在$[0,\delta]$中总是可积的，所以例如对积分 $I_1=\frac{1}{\pi}\int_{0}^{\delta}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}[f(x+t)-f(x+0)]\mathrm{d}x$ 应用第二积分中值定理，知此积分

$$I_1=\frac{1}{\pi}[f(x+\delta)-f(x-0)]\int_{0}^{\xi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}\mathrm{d}t$$
$$=\frac{1}{\pi}[f(x+\delta)-f(x-0)]\int_{0}^{\left(N+\frac{1}{2}\right)\xi}\frac{\sin t}{t}\mathrm{d}t.$$

后面的积分当 $N\to\infty$时极限为$\frac{\pi}{2}$，所以不论 $N$ 和$\delta$ 如何取，它总是有界的. 再由 $f(x)$在$[x,x+\delta]$中为单调的假设(单调函数只有第一类间断点)，对任意 $\varepsilon>0$ 必可找到一个 $\delta(\varepsilon)$使 $|f(x+\delta)-f(x+0)|<\frac{\varepsilon}{4M}$，$M=\sup\limits_{N,\delta}\left|\frac{1}{\pi}\int_{0}^{\left(N+\frac{1}{2}\right)\xi}\frac{\sin t}{t}\mathrm{d}t\right|$，

所以

$$|I_1|<\frac{\varepsilon}{4}.$$

同理当 $\delta$ 充分小时

$$|I_2|=\left|\frac{1}{\pi}\int_{0}^{\delta}\frac{\sin\left(N+\frac{1}{2}\right)}{t}[f(x-t)-f(x-0)]\mathrm{d}t\right|<\frac{\varepsilon}{4}.$$

现在固定这个 $\delta$，它与 $N$ 无关，而(29)中的积分化为

$$\frac{1}{\pi}\int_{\delta}^{\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{t}[f(x+t)+f(x-t)-f(x+0)-f(x-0)]\mathrm{d}t+I_1+I_2$$

对前一积分可以应用黎曼-勒贝格引理而知当 $N$ 充分大时，其绝对值小于 $\varepsilon/2$. 由此知

$$\lim_{N\to\infty}\int_{-\pi}^{\pi}W_N(t)f(x+t)\mathrm{d}t=\lim_{N\to\infty}\frac{2}{\pi}\int_{0}^{\left(N+\frac{1}{2}\right)\pi}\frac{\sin\left(N+\frac{1}{2}\right)t}{2t}[f(x+0)+f(x-0)]\mathrm{d}t$$

$$=\frac{1}{2}[f(x+0)+f(x-0)].$$

但这只是证明了 $W_N(t)$是 $\delta$ 序列，还未证明级数(16)是否在条件$(D)$下收敛于 $f(x)$. 于是我们还需证明

**定理 8**(**狄利克雷定理**)　设 $f(x)$在$[-\pi,\pi]$中适合条件$(D)$，则级数(16)必收敛，其和为

$$\sum_{k=-\infty}^{\infty} c_k e^{ikx}=\frac{1}{2}[f(x+0)+f(x-0)]. \tag{30}$$

**证**　前面已经证明了左方之和为

$$\lim_{N\to\infty}\int_{-\pi}^{\pi} D_N(t)f(x+t)dt,$$

定理 7 则证明了上式右方为$\lim\limits_{N\to\infty}\int_{-\pi}^{\pi} W_N(t)f(x+t)dt$. 所以现在需证的只是

$$\lim_{N\to\infty}\int_{-\pi}^{\pi}[D_N(t)-W_N(t)]f(x+t)dt=0.$$

但是

$$D_N(t)-W_N(t)=\sin\left(N+\frac{1}{2}\right)t\cdot\left[\frac{1}{2\sin\frac{t}{2}}-\frac{1}{t}\right],$$

$\dfrac{1}{2\sin\frac{t}{2}}-\dfrac{1}{t}$在 $t\in[-\pi,\pi]$中只有一个奇点 $t=0$，但是

$$\frac{1}{2\sin\frac{t}{2}}-\frac{1}{t}=\frac{1}{2}\left[\frac{1}{\sin\tau}-\frac{1}{\tau}\right]=\frac{\tau-\sin\tau}{2\tau\sin\tau}$$

$$=\frac{\frac{\tau^3}{3!}-\frac{\tau^5}{5!}+\cdots}{2\tau^2-\frac{2\tau^4}{3!}+\frac{2\tau^6}{5!}-\cdots}=\frac{\tau}{2\cdot 3!}+\sum_{k=2}^{\infty}\alpha_k\tau^k.$$

这里 $\tau=\frac{t}{2}$，而级数在 $\tau=0$ 即 $t=0$ 附近收敛，所以 $D_N(t)-W_N(t)=\Psi(t)\sin\left(N+\frac{1}{2}t\right)$，而 $\Psi$ 在$[-\pi,\pi]$上光滑，因此 $L^1$ 可积. 利用黎曼 – 勒贝格引理即得定理 8 的证明.

上面我们详细地讨论了级数(16)的收敛性问题并且利用了很细密的数学分析方法. 当然读者会问，能不能找到更精密的方法来证明 $D_N(t)$即在连续函数空间中也是一个 $\delta$ 序列呢？答案是否定的，因为早在 1876 年就有人找到了一个连续函数，其傅里叶级数(16)可以在某点不收敛. 甚至有这样的例子，即连续函数的傅里叶级数可以在“许多点”上不收敛. 如果对 $f(x)$只假设可积性，即 $f(x)\in L^1$，科尔莫哥洛夫甚至构造出几乎处处(甚至处处)发散的例子. 如果假设 $f(x)\in L^2$，鲁金在 1915 年就猜测，$L^2$ 函数 $f(x)$的傅里叶级数必几乎处处收敛于 $f(x)$(此级数必在 $L^2$ 意义下收敛则上一章§4 就已讲过了，是一个非常简单而又基本的事实). 这个猜测在到 1966 年才由瑞典数学家卡尔孙(L. Carleson)证明. 总之，我们遇到的是这样一个问题，既刻画一个函数类，又刻画一个级数(16)之类；即刻画出一类$\{c_k\}$，使二者相互一一对应. 由此出现了许多结果，使我们看到傅里叶级数理论是一个内容非常丰富而又十分困难的理论. 但是我们

不能不问，是否能走另外一条路呢？我们想走广义函数的路．下一节我们再详细地讲，现在只简单说几句．

我们现在回到定理 2，并且先作一些不严格的讨论．上面我们已经说了，$\delta(x)$的傅里叶级数是

$$\delta(x)=\sum_{k=-\infty}^{\infty}\frac{1}{2\pi}e^{ikx}.$$

它在 $D'$意义下收敛．如果应用狄利克雷定理，应有

$$\sum_{k=0}^{\infty}\frac{4}{(2k+1)\pi}\sin kx=\begin{cases}1, & 0<x\leqslant\pi,\\ -1, & -\pi\leqslant x\leqslant 0.\end{cases}\tag{31}$$

不过，在间断点 $x=0,\pm\pi$ 应该按狄利克雷定理对级数的和作必要的修改．其实，由 $\delta(x)$的展开式逐项积分也可得到它．这个级数可以与定理 2 的证明联系起来看：若函数本身有第一类间断点，对其傅里叶系数 $c_k=O\left(\frac{1}{k}\right)$．把(31)看成一个波(在信号处理中人们时常这样看，而且把 $x$ 解释为时间)，则(31)是一个“方波”，在信号处理中它是很重要的．其实第四章一开始就有类似结果，见第四章§1 图 4-1-1．

$$\frac{\pi}{4}=\cos x-\frac{\cos 3x}{3}+\frac{\cos 5x}{5}-\cdots.$$

它与(31)形状不同只是因为它们是同一个函数在不同区间上的展开式．

级数(31)在物理上就表明方波可以用不同频率的正弦波之叠加来逼近．由于方波是很有用的，所以物理学家自然想办法造了种种仪器去生成正弦波(这是很容易的)并把它们合成．其实这些仪器都可以看作是专门用于调和分析的计算机．有了现代的计算机以后，这更是不成问题的小事了．本节开始说到的迈克尔孙就设计了一个“计算机”来用(31)的部分和生成方波．可是试用的结果总是不能满意．他生成的波在间断点处总会有些地方超出方波，高度约为半跃度的 9%．这个现象称为吉卜斯(J. W. Gibbs 就是第四章§5 提到的著名的统计物理学家)现象，因为吉卜斯在 1899 年“自然”杂志上致信迈克尔孙指出这是一个一般现象．这个缺点显然是(31)不一致收敛造成的，因为如果(31)的部分和 $S_N(f)$在$[-\pi,\pi]$上一致收敛，则其极限函数一定连续，而现在极限函数显然是不连续的．所谓一致收敛性就是说在 $f(x)$的曲线上下各划出一个宽为 $\varepsilon$ 的带形，当 $N>N(\varepsilon)$时，$S_N(f)$的曲线一定在$[-\pi,\pi]$区间中全部位于此带内．吉卜斯现象则指出，只要 $\varepsilon<$半跃度的 9%则不论 $N$ 取多大，$S_N(f)$的曲线在间断点附近一定会“冒出”这个带形之外．当我们讨论不一致收敛时，注意到在几何上什么是不一致收敛是很有必要的．从实际上生成方波来说，克服吉卜斯现象以改善收敛性是很有必要的．为此，我们可以采用无穷级数的求和法以改善收敛性．

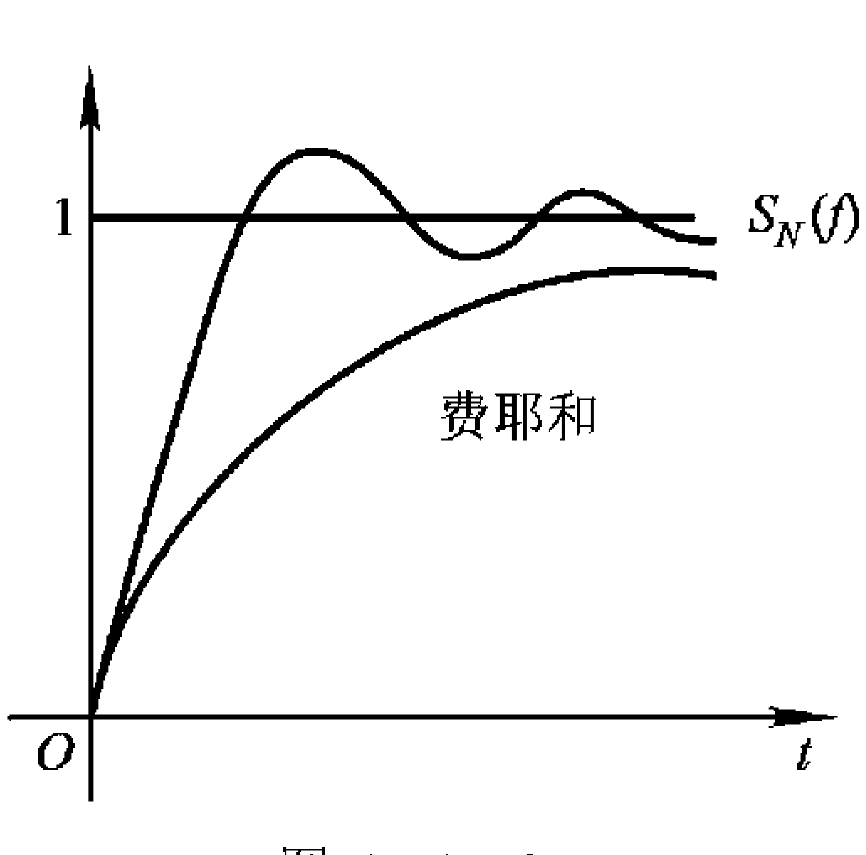



图 5-1-2

无穷级数理论中有许多求和法，我们下面要讲的是所谓 $C$—1 求和法($C$ 是指意大利数学家切萨罗(Cesàro)，把它应用于傅里叶级数的则是匈牙利数学家费耶(Fejer))．其法如下：设有级数

$\sum\limits_{k=0}^{\infty} u_k$，记其部分和为 $U_n = \sum\limits_{k=0}^{n} u_k$，我们考虑序列 $U_0, U_1, \cdots, U_n$ 的算术平均值并记为 $\sigma_n = \frac{1}{n+1}(U_0 + U_1 + \cdots + U_n)$，现在问 $\sigma_n$ 之极限是什么？如果 $U_n \to U$，用一点比较简单的极限计算可知 $\sigma_n \to U$，但其逆并不成立．所以$\{\sigma_n\}$的收敛性“好”于$\{U_n\}$，若 $\sigma_n \to U$，则说级数 $\sum\limits_{k=0}^{\infty} u_k$ 为 $C$—1 可求和(summable)，而其 $C$—1 和为 $U$．现在把它应用于级数(16)，前面我们已多次计算过它的部分和可写为

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(u) \sum_{k=-N}^{N} \mathrm{e}^{\mathrm{i}k(u-x)} \mathrm{d}u = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x+t) \sum_{k=-N}^{N} \mathrm{e}^{\mathrm{i}kt} \mathrm{d}t.$$

我们说这里的$\sum$是它的部分和 $S_N$，尽管实际上它是 $2N+1$ 项之和．计算它们的算术平均值会得到

$$K_N(t) = \sum_{k=-N}^{N} \left(1 - \frac{|k|}{N+1}\right) \mathrm{e}^{\mathrm{i}kt}.$$

$K_N(t)$称为费耶核，可以证明

$$K_N(t) = \frac{1}{N+1}\left(\frac{\sin \frac{N+1}{2}t}{\sin \frac{t}{2}}\right)^2.$$

证法要点如下：

注意到

$$\sin^2 \frac{t}{2} = \frac{1}{2}(1 - \cos t) = -\frac{1}{4}\mathrm{e}^{-\mathrm{i}t} + \frac{1}{2} - \frac{1}{4}\mathrm{e}^{\mathrm{i}t},$$

所以，经过简单计算即有

$$\begin{aligned} \sin^2 \frac{t}{2} \cdot K_N(t) &= \left(-\frac{1}{4}\mathrm{e}^{-\mathrm{i}t} + \frac{1}{2} - \frac{1}{4}\mathrm{e}^{\mathrm{i}t}\right) \sum_{k=-N}^{N} \left(1 - \frac{|k|}{N+1}\right)\mathrm{e}^{\mathrm{i}kt} \\ &= \frac{1}{N+1}\left(-\frac{1}{4}\mathrm{e}^{-\mathrm{i}(N+1)t/2} + \frac{1}{2} - \frac{1}{4}\mathrm{e}^{\mathrm{i}(N+1)t/2}\right) \\ &= \sin^2 \frac{N+1}{2}t. \end{aligned}$$

值得注意的是，$K_N(t)$即使在连续函数空间中也是 $\delta$ 序列，实际上，对任何$[-\pi,\pi]$上的连续函数 $f(x)$

$$\lim_{N\to\infty}\int_{-\pi}^{\pi} f(u) K_N(u-x) \mathrm{d}u = f(x).$$

这里的证明不难，我们都略去了．总之，连续函数 $f(x)$的傅里叶级数(16)必为 $C$—1 可求和而以 $f(x)$为其 $C$—1 和．看一下图 4-1-2 中标注了费耶和的曲线，就看得出它克服了吉卜斯现象．

## § 2 傅里叶变换

**1．$L^1(\mathbf{R}^n)$函数的傅里叶变换** 傅里叶级数出现后不久，就有了傅里叶积分——傅里叶变

换．前者把$[-\pi,\pi]$上的周期函数 $f(x)$“写成”$\sum_{k=-\infty}^{\infty} c_k e^{ikx}$，或者也可以说是把一个函数“变成”一个序列$\{\cdots,c_{-2},c_{-1},c_0,c_1,c_2,\cdots\}$所以它是函数空间到序列空间的一个映射．也可以说是一个物理过程有了两个不同的“表示”：一是表为 $x$(可以是时间，所以下面改写成 $t$)的函数，一是表为频率 $k$ 的函数，不过时间 $t$ 是取连续值的，而频率是取整数值的(因为我们使用了指数函数表示正弦波，所以这些整数也可以是负的)．但是一个自然现象很少是真正周期的．例如上面我们说 $f(x)$是周期函数，其实它原来可以定义在 $\mathbf{R}$ 上，只不过我们取了它的在$[-\pi,\pi]$上的一段再作周期延拓，这样才得到周期函数的．由上节的例，在有限区间上的 S－L 问题又时常是有离散谱的，这是把周期函数 $f(x)$在一个周期内展开为固有函数的级数，就是傅里叶级数——的问题的来源．因此自然要问，如果考虑的是定义在一个无穷区间上的函数，那么会得到什么？事实上当年傅里叶就是从研究一个半无限长的杆上的温度分布得出傅里叶积分的．具体情况下如：设在杆$[0,+\infty]$，上有温度 $u(x,t)$分布，由第三章第一节知，$u(x,t)$满足热传导方程

$$\frac{1}{a^2}\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}. \tag{1}$$

我们设杆上初始时刻的温度分布是已知的，而在杆的左端 $x=0$ 处温度又保持在 $u(0,t)=0$，于是除方程(1)以外还有边值条件和初值条件

$$u(0,t)=0,\quad u(x,0)=\varphi(x)\quad(\text{已知函数}). \tag{2}$$

可以仿照上一节求解弦振动方程的办法来解它，令 $u(x,t)=X(x)T(t)$代入方程(1)，化简后有

$$T'(t)/a^2T(t)=X''(x)/X(x)=\lambda.$$

于是和前面一样，试由它以及边值条件 $X(0)=0$ 求出 $X(x)$，我们仍设 $\lambda=\mu^2>0$，于是

$$X(x)=Ae^{\mu x}+Be^{-\mu x}.$$

但是，若 $A\neq 0$，则当 $x\to+\infty$时会有 $X(x)\to\infty$．这在物理上是说不通的．因此，必须有 $A=0$．再由 $X(0)=0$，又得 $X(0)=Be^0=B=0$，于是又归结为一个平凡解．这样，我们否定了 $\lambda=\mu^2$ 的可能性．这里有一点十分值得注意：物理上合理性的要求限制了函数在无穷远处的性状．而后者又起了边值条件的作用．这种限制可以是例如 $X(x)$在 $x\to+\infty$时为 $O(1/x^a)$，可以是 $X(x)\in L^1([0,+\infty))$或 $X(x)\in L^2([0,+\infty))$等等．所以我们看到，限制在某个函数空间的框架内研究某个数学问题不只是有数学上的意义，而且常有物理上的意义．我们暂时只能讲到这里为止．总之，$\lambda=\mu^2$ 已被排除．同样，我们也会排除 $\lambda=0$，而余下的只有 $\lambda=-\mu^2$，从而

$$X(x)=A\cos\mu x+B\sin\mu x.$$

由 $X(0)=0$ 可得

$$X(x)=\sin\mu x, \tag{3}$$

我们略去了常数(系数 $B$ 取为 1)，但是我们不能说一切 $\mu\in\mathbf{R}$ 都是固有值而(3)为其固有函数．我们只说以下的边值问题

$$\begin{aligned}&X''(x)-\lambda X(x)=0,\qquad x\in[0,+\infty),\\&X(0)=0,\quad X\text{ 在 }x\to+\infty\text{ 时不趋于}\infty,\end{aligned} \tag{4}$$

(最后一个要求也可以是 $X(x)\in L^1([0,+\infty))$等等，现在只是作一个例子来提出问题)的“谱”是 $\lambda=-\mu^2(\mu\in\mathbf{R})$．什么时候说固有值，什么时候只能说“谱”，这是属于谱论的问题．我们现在

只说问题(4)有“连续谱”. 相应于 $\lambda=-\mu^2$,有

$$T'(t)+a^2\mu^2T(t)=0,$$

所以

$$T(t)=\Phi(\mu)e^{-a^2\mu^2t},\quad \mu\in\mathbf{R},t>0,$$

而问题(1),(2)之解现在不是对 $\mu$ 求和,而是对 $\mu$ 求积分:

$$u(x,t)=\int_{-\infty}^{\infty}\Phi(\mu)e^{-a^2\mu^2t}\sin\mu x\mathrm{d}\mu.$$

问题就在于如何选择 $\Phi(\mu)$,使当 $t=0$ 时得到

$$\psi(x)=\int_{-\infty}^{+\infty}\Phi(\mu)\sin\mu x\mathrm{d}\mu. \tag{5}$$

自傅里叶以后,以上形状的积分(还有将正弦改为余弦所得的积分)通称为傅里叶积分. 但是我们还要从更广泛的视野来看它. 首先我们引入复记号,这样写出一个与(5)相应的式子

$$\varphi(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi(\mu)e^{i\mu x}\mathrm{d}\mu \tag{6}$$

积分号前多了一个因子$\frac{1}{2\pi}$自然无关紧要,双方乘以 $e^{-i\xi x}$,并对 $x$ 求积分,如果我们允许任意地交换积分次序将有

$$\int_{-\infty}^{+\infty}\varphi(x)e^{-i\xi x}\mathrm{d}x=\int_{-\infty}^{\infty}\Phi(\mu)\mathrm{d}\mu\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i(\mu-\xi)x}\mathrm{d}x. \tag{7}$$

我们提出:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{i(\mu-\xi)}\mathrm{d}x=\delta(\mu-\xi). \tag{8}$$

这个式子看起来令人吃惊,但是很有道理. 如果把积分写成和,把 $x$ 变成 $\mathbf{R}$(由 $-\infty$ 到 $+\infty$)的分点 $x=k$,$k$ 为整数,则上式左方成为

$$\sum_{k=-\infty}^{\infty}\frac{1}{2\pi}e^{i(\mu-\xi)k}. \tag{9}$$

但是上一节(27)式中我们讲过 $\delta(x)$之傅里叶级数正是

$$\delta(x)=\sum_{h=-\infty}^{\infty}\frac{1}{2\pi}e^{ikx}, \tag{10}$$

那是离散谱的情况. 现在换成了连续谱,级数应该变成积分. 在(10)中把 $x$ 写成$(\mu-\xi)$,注意到 $\Delta x=(k+1)-k=1$,(10)式立即变成(8)式,于是

$$\int_{-\infty}^{\infty}\varphi(x)e^{-i\xi x}\mathrm{d}x=\int_{-\infty}^{\infty}\Phi(\mu)\delta(\mu-\xi)\mathrm{d}\mu=\Phi(\xi). \tag{11}$$

这样我们得到一对公式(6)和(11).

上面的证明当然算不得证明,但是它确实是物理学家考虑问题的办法,而且它揭示了十分重要的思想. 我们以为(6)和(11)是一对变换和逆变换,(11)式把函数 $\varphi(x)$映到 $\Phi(\xi)$,而(6)式就是它的逆. 如果我们把 $\varphi(x)$看作一个信号,$x$ 表示时间,则(11)表示它是由各种不同频率 $\xi$ 的谐波 $e^{i\xi x}$构成的,$\Phi(\xi)$就是在谐波构成 $\varphi(x)$这个信号时,谐波 $e^{i\xi x}$ 占了多少分量. 所以(11)式是一个谐波分析的公式. 把它们叠加起来,(6)则是表示这些谐波的合成的公式. 但是不论是 $\varphi(x)$或 $\Phi(\xi)$都表示同一个信号,前者我们称为该信号在时域的表示,后者则是它在频域中的

表示．更重要的是在量子力学中它们含意还更深刻．这时 $\varphi(x)$ 是波函数，它是用粒子的位置 $x$ 来表示的．在量子力学中 $\xi$ 则表示粒子的动量．$\Phi(\xi)$ 则是波函数用粒子的动量来表示的．但是它们都是同一个状态的不同表示(物理学中称为“表象”)．前者是 $x$ 表象，后者是 $\xi$ 表象．但是在经典物理学中决定一个粒子的状态应同时了解其位置 $x$ 与动量 $\xi=mv$ (所以也就是速度)，二者互相独立．现在位置 $x$ 与动量 $\xi$ 之间是“有联系的”．迄今为止，人们可能会有一个错觉，以为两个变量之间的“联系”大概总是一种函数关系．可是在量子力学中却不能说一个粒子的位置与动量间有函数关系．二者之间的关系就是著名的“测不准关系”①．对前者的测量会影响到后者的测量．反过来也是一样．这在经典物理学中是不可想像的．而这种测不准关系正是(6)与(11)这样一对变换与逆变换的必然推论．如此说来，信号的时域与频域表示之间也有(6)与(11)这样的“联系”，那么二者之间是否也有某种测不准关系呢？结论是肯定的．从物理学角度看，“测不准原理”是微观世界的物理学——量子物理与宏观世界的物理学——经典物理的分水岭．但是从数学角度来看则是：某一类函数空间中，只要有(6)和(11)这样的变换关系就一定有测不准关系．不但有量子的测不准关系，也有经典的测不准关系．因此，(6)与(11)这样的变换关系是十分深刻的．我们称(11)是某一函数空间中的 $\varphi(x)$ 到另一函数空间中的 $\Phi(\xi)$ 的傅里叶变换，而(6)则称为其逆傅里叶变换．在下面我们规定采用以下的记号，记 $\Phi(\xi)=\hat{\varphi}(\xi)$ 而(6)式的逆变换则记为 $(F^{-1}\Phi)(x)=f(x)$ (有的书上则记 $\hat{\varphi}(\xi)$ 为 $F(\varphi)(\xi)$)．(6)与(11)形状不太对称，因为(6)式中多了一个因子 $\frac{1}{2\pi}$．为了对称起见，许多文献上常把它们写作

$$\Phi(\xi)=\int_{-\infty}^{\infty}\varphi(x)\mathrm{e}^{-2\pi\mathrm{i}\xi x}\mathrm{d}x,$$

$$\varphi(x)=\int_{-\infty}^{\infty}\Phi(\xi)\mathrm{e}^{2\pi\mathrm{i}\xi x}\mathrm{d}\xi.$$

这相当于把频率换成了圆频率．

我们下面的任务就是讨论这种变换的性质，特别是给它们以严格的证明(由此也就给(8)式以明确的含意)．但是这时首先要问，在什么框架下讨论它们，也就是上面说的“某一函数空间”中的 $\varphi(x)$，究竟是指哪一个函数空间．

首选的空间自然是 $L^1$．因为(6)式中出现了积分，而从傅里叶级数的情况来看黎曼积分是不够用的．下面我们不限于 $L^1(\mathbf{R}^1)$，而用一般的 $L^1(\mathbf{R}^n)$，因此我们给出

**定义 1** 设 $f(x)\in L^1(\mathbf{R}^n)$，我们称

$$\hat{f}(\xi)=\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x \tag{12}$$

为其傅里叶变换，这里 $x\xi=\sum_{j=1}^{n}x_j\xi_j$ 时常记作 $x\cdot\xi$，$\langle x,\xi\rangle$ 或 $\langle\xi,x\rangle$．

但我们暂时还未宣称

$$f(x)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\hat{f}(\xi)\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi$$

① 全国自然科学名词审定委员会 1988 公布的《物理学名词》上，改称“不确定[度]关系”．下文提到的“测不准原理”在《物理学名词》上改称“不确定[性]原理”．

为逆傅里叶变换,因为 $f(x)\in L^1(\mathbf{R}^n)$时,$\hat{f}(\xi)$在什么空间中并不清楚,所以这个积分暂时也没有定义.

给出了这个定义后,我们要讨论它的种种性质,其中最重要的是它关于 $x$ 的线性变换的性质以及关于$f(x)$之代数与分析运算的性质. 其中有一个是在上一节已经证明的定理 3,即

**黎曼-勒贝格引理** 若 $f(x)\in L^1(\mathbf{R}^n)$则

$$\lim_{\xi\to\infty}\hat{f}(\xi)=0.$$

这个定理下面我们就只直接使用而不再证明了.

但是,从物理学的角度看来,$L^1(\mathbf{R}^n)$是不合用的.因为例如信号处理问题中,$|\hat{f}(\xi)|^2\mathrm{d}\xi$ 代表频率在 $\xi$ 到 $\xi+\mathrm{d}\xi$ 中的谐波成分之功率. 因为一个信号的功率总是有限的,所以必须有

$$\int_{\mathbf{R}^n}|\hat{f}(\xi)|^2\mathrm{d}\xi<+\infty.$$

再看量子力学的情况. 这时 $\varphi(x)$作为波函数其物理解释是:$|\varphi(x)|^2\mathrm{d}x$ 是粒子位于 $x$ 到 $x+\mathrm{d}x$ 之间的概率. 既然是概率,则必有

$$\int_{\mathbf{R}^n}|\varphi(x)|^2\mathrm{d}x=1.$$

因此,从物理学的角度来看,(6)和(11)应该看成是由 $L^2(\mathbf{R}^n)$到 $L^2(\mathbf{R}^n)$的映射. 上一章中我们指出 $L^2(\mathbf{R}^n)$有特别重要的结构,因此这时我们除了要再点明 $L^1(\mathbf{R}^n)$中傅里叶变换的性质外,更要从 $L^2(\mathbf{R}^n)$特有的结构来考察它.

然后,我们又要在广义函数框架下去看傅里叶变换. 到这里我们才最后完成(6)与(11)的证明,到那时(8)式的意义也就明确了. 我们得到了一个比较完整的理论. 其实广义函数理论的出现主要理由之一正是为了傅里叶变换理论的需要.

现在我们在 $L^1(\mathbf{R}^n)$中讨论傅里叶变换.

**定理 1** 若 $f(x)\in L^1(\mathbf{R}^n)$,则$\hat{f}(\xi)$必在 $\mathbf{R}^n$ 上有界、一致连续,而且在无穷远处为 0.

**证** 易见

$$|\hat{f}(\xi)|\leqslant\int_{\mathbf{R}^n}|f(x)\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x|=\int_{\mathbf{R}^n}|f(x)|\mathrm{d}x=\|f\|_{L^1},$$

所以$\hat{f}(\xi)$之有界性自明. 又

$$\begin{aligned}|\hat{f}(\xi+h)-\hat{f}(\xi)|&\leqslant\int_{\mathbf{R}^n}|f(x)||\mathrm{e}^{-\mathrm{i}xh}-1|\mathrm{d}x\\&=2\int_{\mathbf{R}^n}|f(x)|\left|\sin\frac{xh}{2}\right|\mathrm{d}x=I_1+I_2.\end{aligned}$$

我们将此积分分为两部分,其一是

$$I_1=2\int_{|x|>R}\left|f(x)\sin\frac{xh}{2}\right|\mathrm{d}x\leqslant2\int_{|x|>R}|f(x)|\mathrm{d}x,$$

故对任意 $\varepsilon>0$,必可找到一个与 $x,h$ 均无关的$R$,使上面的积分小于 $\varepsilon/2$,由此固定 $R$.再看

$$I_2=2\int_{|x|<R}\left|f(x)\sin\frac{xh}{2}\right|\mathrm{d}x.$$

因为$\left|\sin\frac{xh}{2}\right|\leqslant\frac{1}{2}|xh|\leqslant\frac{R}{2}\|h\|$（注意这里 $xh$ 是内积，故$|xh|\leqslant\|x\|\cdot\|h\|\leqslant R\|h\|$，$\|h\|^2=\sum_{i=1}^{n}h_i^2$. 但以下 $n$ 维向量的长仍常以$|\cdot|$表示），故

$$|I_2|\leqslant R\int_{|x|<R}|f(x)|\mathrm{d}x\cdot\|h\|<\frac{\varepsilon}{2}.$$

只要$\|h\|$充分小即可（注意现在 $R$ 是固定的）. 所以，当$|\xi|\to\infty$时$\hat{f}(\xi)\to 0$ 就是黎曼－勒贝格引理.

但是$\hat{f}(\xi)$一般并不可积. 例如设

$$f(x)=\begin{cases}1, & |x|<1,\\ 0, & |x|>1.\end{cases}$$

$f(x)\in L^1(\mathbf{R}^1)$（我们没有规定$|x|=1$ 时 $f(x)$之定义，但这没有关系，因为 $L^1(\mathbf{R}^1)$函数只要几乎处处有定义即可），但是它的傅里叶变换是

$$\hat{f}(\xi)=\int_{-\infty}^{\infty}f(x)\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x=\int_{-1}^{1}\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x=-\frac{\mathrm{e}^{-\mathrm{i}x\xi}-\mathrm{e}^{\mathrm{i}x\xi}}{\mathrm{i}\xi}=\frac{2\sin\xi}{\xi}. \tag{13}$$

它虽然连续，而且当$|\xi|\to\infty$时趋向 0，但它并不属于 $L^1(\mathbf{R}^1)$. 因为勒贝格可积一定是指绝对可积，而这个函数却不是绝对可积的. 因此，对这个 $f(x)$，(6)式不可能成立. 由 $f(x)$到$\hat{f}(\xi)$的变换称为傅里叶变换，而当 $f(x)\in L^1$ 时，它可以用(11)表示. 那么由$\hat{f}(\xi)$到 $f(x)$就应称为逆傅里叶变换. 上面我们用形式的办法"推导"出来它应该写为(6)式，但实际上，既然连$\hat{f}(\xi)\in L^1$都不能保证，(6)式又有什么意义呢？上一节讲傅里叶级数时，先给出了 $f(x)\in L^1([-\pi,\pi])$，然后就可以算出它的傅里叶系数 $c_k=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)\mathrm{e}^{-\mathrm{i}kx}\mathrm{d}x$. 傅里叶系数序列$\{c_k\}$（看成是定义在整数集 $\mathbf{Z}$ 上的函数）就相当于$\hat{f}(\xi)$. 在上一节，我们开始先说 $f(x)$的傅里叶级数 $f(x)\sim\sum_{k=-\infty}^{\infty}c_k\mathrm{e}^{\mathrm{i}kx}$ 是形式级数，并因此使用了记号～. 其实现在的逆变换公式(6)也只是一个"形式的"对象. 上一节我们用了很大力量去讨论傅里叶级数的收敛问题，其实还只接触到这个重大问题比较浅层的结果，也就是说，$L^1$ 函数 $f(x)$之傅里叶级数收敛于 $f(x)$即由$\{c_k\}$可以通过$\sum_{k=-\infty}^{\infty}c_k\mathrm{e}^{\mathrm{i}kx}$恢复$f(x)$的必要充分条件没有找到. 也就因此，才产生了诸如费耶求和等各种方法以恢复 $f(x)$的问题. 现在的问题也是一样，如果我们记 $\mathbf{R}^n$ 上连续而无穷远处为 0 的函数之空间为 $C_0(\mathbf{R}^n)$，则定理 1 告诉我们（用 $F$ 表示傅里叶变换，并看作一个映射）$F: L^1(\mathbf{R}^n)\to C_0(\mathbf{R}^n)$，而且 $F$ 的"像"其实是 $C_0(\mathbf{R}^n)$的一个真子空间，而数学家们迄今未能完全刻画它. 所以问题在于，如何找到一个空间 $T$，它要足够宽广，包括 $L^1$，$L^2$ 等等；在 $T$ 上又可定义傅里叶变换$F$，使 $F:T\to T$ 是一个双方单值的映射，广义函数理论很大程度上正是为了回答这个问题而出现的.

下面来看当 $f(x)$ 变化时，$\hat{f}(\xi)$ 怎样随之改变. 一个明显的结果是 $f(x)\to\hat{f}(\xi)$ 是线性映射，若 $f_i(x)\in L^1(\mathbf{R}^n)$，$c_i$ 是复常数，$i=1,2$，则

$$\widehat{(c_1 f_1 + c_2 f_2)}(\xi) = c_1\hat{f}_1(\xi) + c_2\hat{f}_2(\xi). \tag{14}$$

$L^1(\mathbf{R}^n)$ 函数的乘积一般不再在 $L^1(\mathbf{R}^n)$ 中，但是有一种运算称为卷积，其形式的定义是

$$(f * g)(x) = \int_{\mathbf{R}^n} f(t)g(x-t)\mathrm{d}t = \int_{\mathbf{R}^n} f(x-t)g(t)\mathrm{d}t. \tag{15}$$

首先的问题是，当 $f,g\in L^1(\mathbf{R}^n)$ 时，上面的积分是否仍是 $L^1(\mathbf{R}^n)$ 函数. 当 $f,g$ 为黎曼可积时，上述积分仍是一个黎曼可积函数，这可以用富比尼定理来证明，对于 $L^1(\mathbf{R}^n)$ 函数，富比尼定理我们已在第四章中讲到，现在再把结果陈述如下：

**定理 2（富比尼（Fubini）定理）**　若 $f(x,y)$ 是 $L^1(\mathbf{R}^n\times\mathbf{R}^n)$ 函数，则对几乎所有 $x$（或 $y$），$f(x,y)\in L(\mathbf{R}_y^n)$（或 $f(x,y)\in L^1(\mathbf{R}_x^n)$），而且 $\int_{\mathbf{R}^n} f(x,y)\mathrm{d}y\in L^1(\mathbf{R}_x^n)$（或 $\int_{\mathbf{R}^n} f(x,y)\mathrm{d}x\in L^1(\mathbf{R}_y^n)$），且

$$\begin{aligned}\iint_{\mathbf{R}^n\times\mathbf{R}^n} f(x,y)\mathrm{d}x\mathrm{d}y &= \int_{\mathbf{R}^n}\mathrm{d}x\int_{\mathbf{R}^n} f(x,y)\mathrm{d}y \\ &= \int_{\mathbf{R}^n}\mathrm{d}y\int_{\mathbf{R}^n} f(x,y)\mathrm{d}x.\end{aligned} \tag{16}$$

**托乃利（Tonelli）定理**　若 $f(x,y)\geqslant 0$ 为可测函数，则若（16）式中的三个积分中只要有一个存在，另外两个一定也存在而且（16）式成立.

现在把这个定理用于卷积（15），为此，我们考虑 $|f(x-t)g(t)|$. 显然 $\int_{\mathbf{R}_x^n}|f(x-t)g(t)|\mathrm{d}x = |g(t)|\int_{\mathbf{R}^n}|f(x-t)|\mathrm{d}x = |g(t)|\int_{\mathbf{R}_x^n}|f(x)|\mathrm{d}x\in L(\mathbf{R}_t^n)$. 因此，$\int_{\mathbf{R}_t^n}|g(t)|\mathrm{d}t\int_{\mathbf{R}_x^n}|f(x-t)|\mathrm{d}x < +\infty$，由托乃利定理即知 $\iint_{\mathbf{R}^n}|f(x-t)g(t)|\mathrm{d}x\mathrm{d}t < +\infty$，即 $f(x-t)g(t)\in L^1(\mathbf{R}_t^n\times\mathbf{R}_x^n)$，应用托乃利定理，即知（15）所定义的 $(f*g)(x)\in L^1(\mathbf{R}^n)$，而且重复前证的步骤即有

$$\left|\int_{\mathbf{R}_x^n}(f*g)(x)\mathrm{d}x\right| \leqslant \int_{\mathbf{R}_t^n}|g(t)|\mathrm{d}t\cdot\int_{\mathbf{R}_x^n}|f(x)|\mathrm{d}x, \tag{17}$$

或写作

$$\|f*g\|_{L^1}\leqslant\|f\|_{L^1}\cdot\|g\|_{L^1}. \tag{17$'$}$$

卷积很像是一个乘法. 它适合交换律（这一点由定义（15）式即知）、结合律

$$(f*g)*h = f*(g*h)$$

和分配律

$$(f_1+f_2)*g = f_1*g + f_2*g.$$

所以 $L^1(\mathbf{R}^n)$ 不但是一个线性空间，而且其中还有适合以上运算律的"乘法"，这就称为一个"代数". 再考虑到其中的范数适合不等式（17），就称为一个巴拿赫代数. 但是这个乘法又与"普通"

的乘法有些不同,例如,其中没有单位元1(即与任一 $L^1$ 函数 $f$ 之卷积均等于 $f$ 本身的 $L^1$ 函数"1":$f*$"1"$=f$). $\delta$ 函数适合此式,但 $\delta$ 又不是一个 $L^1$ 函数. 又例如有些乘法是没有零因子的,即由 $f\cdot g=0$ 必有 $f=0$ 或 $g=0$. 卷积就没有零因子.这就是梯其马希(Titchmarsh)定理:若 $f,g\in L^1(\mathbf{R}^n)$ 而且 $f*g=0$,则 $f,g$ 中至少有一个几乎处处为0,但这并不容易证明.

卷积是很重要的运算. 我们实际上早已见过它了,例如磨光运算就是函数与磨光核的卷积:$\delta$ 序列 $K_n(x)\to\delta(x)$ in $D'$ 就是 $f*K_n\to f$ 对一切 $f\in D$ 成立(由此可以想到 $\delta$ 函数是卷积的单位元.即上述的"1"),柯西积分 $\frac{1}{2\pi i}\int_L\frac{f(\zeta)}{\zeta-z}d\zeta$ 即 $f$ 与 $\frac{1}{2\pi i}\frac{1}{\zeta}$ 的卷积(复的). 不过我们现在最关心的是卷积与傅里叶变换的关系.

**定理3** 若 $f,g\in L^1(\mathbf{R}^n)$,则其卷积 $h(x)=(f*g)(x)$ 之傅里叶变换等于 $f$ 与 $g$ 之傅里叶变换之积:

$$\hat{h}(\xi)=\hat{f}(\xi)\cdot\hat{g}(\xi).$$

**证** 应用富比尼定理即有

$$\begin{aligned}\hat{h}(\xi)&=\int_{\mathbf{R}^n}h(x)e^{-ix\xi}dx=\int_{\mathbf{R}^n}dx\int_{\mathbf{R}^n}f(t)g(x-t)e^{-i\xi x}dt\\&=\int_{\mathbf{R}^n}dx\int_{\mathbf{R}^n}f(t)e^{-i\xi t}\cdot g(x-t)e^{-i\xi(x-t)}dt\\&=\int_{\mathbf{R}^n}f(t)e^{-i\xi t}dt\int_{\mathbf{R}^n}g(x)e^{-i\xi x}dx\\&=\hat{f}(\xi)\cdot\hat{g}(\xi).\end{aligned}$$

对于 $\hat{f}$ 与 $\hat{g}$ 之卷积其傅里叶逆变换也应有类似公式. 但是现在连 $\hat{f}*\hat{g}$ 暂时都无法定义,所以只好在后面用广义函数去讨论.

这个定理不但在数学理论中有重要意义,而且在实际工作中很有用. 设有一个信号 $f(t)$(我们使用它的时域表示),我们已经明白,它是由各种不同频率的谐波合成的. 所谓滤波就是希望从这些成分中删除某些成分. 为了从数学上考查这个过程,我们转到此信号的频域表示,即 $\hat{f}(\xi),\xi\in\mathbf{R}^1$. 现在我们要求删除其中的高频成分,例如删除 $|\xi|\geqslant R_0$ 的一切高频成分,能完成这个任务的设备就叫做低通滤波器. 为此,我们考虑一个函数

$$F(\xi)=\begin{cases}1 & |\xi|<R_0,\\0 & |\xi|\geqslant R_0,\end{cases}\tag{18}$$

并考虑新的信号 $F(\xi)\hat{f}(\xi)$,它就是从原来信号中删除了高频成分的结果. 如果 $F(\xi)=\hat{w}(\xi)$,则

$$F(\xi)\hat{f}(\xi)=\widehat{w*f}(\xi).$$

因此回到时域即知

$$(w*f)(t)=\int_{-\infty}^{\infty}f(t)w(x-t)dt.$$

实际上，如果逆变换公式(6)成立的话

$$
\begin{aligned}
w(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\xi)e^{i\xi t}d\xi \\
&= \frac{1}{2\pi}\int_{-R_0}^{R_0} F(\xi)e^{i\xi t}d\xi = \frac{e^{iR_0 t} - e^{-iR_0 t}}{2\pi i t} \\
&= \frac{\sin R_0 t}{\pi t}
\end{aligned}
$$

就是(18)的时域表示．而从数学上看，上述滤波器就是一个完成卷积运算

$$
(w * f)(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(t)\frac{\sin R_0(t-\tau)}{t-\tau}d\tau \tag{19}
$$

的装置．所以函数$\frac{\sin R_0 t}{\pi t}$可以称为这个滤波器的仪器函数．

(19)也称为“窗式傅里叶变换(window Fourier transtorm)．

不过我们要注意上面我们并没有规定信号的时域表示 $f(t)\in L^1(\mathbf{R}_t)$，而 $F(\xi)$上面已说过了不一定是 $L^1(\mathbf{R}_\xi)$函数，因此即使逆变换公式(6)对 $L^1(\mathbf{R}_t)$函数得到了证明，也不能应用于此．所有这一切都表明了完全有必要在广义函数框架下讨论傅里叶变换．此外，在上一节中我们正是针对“窗口函数”(18)提出了吉卜斯现象．同样，在窗式傅里叶变换(19)中也一定会出现这种情况．正如在傅里叶级数时我们采用费耶核$\left[\frac{\sin\left(N+\frac{1}{2}\right)t}{\sin\frac{t}{2}}\right]^2$来代替狄利克雷核，而且这是通过将部分和 $S_N$ 改为$\sigma_N = \sum_{j=-N}^{N}\left(1-\frac{|j|}{N+1}\right)u_j$来实现的．现在也一样，我们只要把窗口函数改为$F(\xi)(1-|\xi|/R)$，从而把(19)改为$(\sin R_0 t/\pi R_0 t)^2 * f$也能实现同样的目的．

现在回到傅里叶变换最引人注目的性质，即微分运算在傅里叶变换下成为什么．我们先形式地看一下会有什么结果，然后再去证明它．为简单计只看 $\mathbf{R}^1$ 的情况．

由(12)

$$
\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-ix\xi}dx,
$$

双方对 $\xi$ 求导即得

$$
\frac{d}{d\xi}\hat{f}(\xi) = \frac{1}{i}\int_{-\infty}^{\infty} xf(x)e^{-ix\xi}dx = \frac{1}{i}(\widehat{xf})(\xi). \tag{20}
$$

再由(6)

$$
f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\xi)e^{ix\xi}d\xi,
$$

双方对 $x$ 求导，有

$$
f'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} i\xi\hat{f}(\xi)e^{ix\xi}d\xi,
$$

注意到(6)式实际上就是逆变换的公式，仿此我们可以想到应有

$$
i\xi\hat{f}(\xi) = \int_{-\infty}^{\infty} f'(x)e^{-ix\xi}dx. \tag{21}
$$

(20)与(21)表示,对原来的函数"乘以自变量"(或对自变量求导),导致对傅里叶变换"对新的自变量求导"(或乘以"自变量"),反过来也是一样.(但还应该适当地加上因子±i.)这是一种新对偶关系.我们就简单地说:同一物理过程的 $x$ 表象乘以自变量(或求导)就得出该过程的 $\xi$ 表象的求导(或乘以自变量).更简单地说就说在傅里叶变换下 $\frac{1}{\mathrm{i}}\partial_x$ 与 $\xi$(乘以 $\xi$)是对偶的,$-\frac{1}{\mathrm{i}}\partial_\xi$ 与 $x$ 是对偶的.这个论断在量子物理中有极大的意义,因为量子物理与经典物理有一个根本区别,即在经典物理中,物理量用函数来表示,而在量子物理中,物理量用算子来表示(当然究竟是作用在什么空间上的何种算子需要详加说明),而傅里叶变换实现了二者之间的对应.

放下物理的解释来看数学,就发现上面的"推导"确有许多问题.例如它要用到 $f(x)$ 的积分,则应该有 $f(x)\in L^1$,同样还应有 $xf(x)\in L^1$,$\hat{f}(\xi)$,$\frac{\mathrm{d}}{\mathrm{d}\xi}\hat{f}(\xi)\in L^1$ 等等.通常,要求一个函数属于 $L^1$ 可以归结为对它在无穷远处衰减速度的限制.例如,若 $\hat{f}(\xi)$ 当 $\xi\to\infty$ 时为 $O(\xi^{-2-\varepsilon})$,则 $f(x)$ 可求导.于是我们又看到对 $\hat{f}(\xi)$ 增长速度(即是衰减速度)的限制导致了 $f(x)$ 的某种程度的可微性.同样,$f(x)$ 在无穷远处的增长速度决定了 $\hat{f}(\xi)$ 的可微性.这个现象我们已在上一节讲到傅里叶级数的收敛性时强调过其重要性.在广义函数框架下处理傅里叶变换也是以此为基础的.现在我们正式提出和证明相关的定理.

**定理 4** (i)设 $f(x)\in L^1(\mathbf{R}^n)$,$x_jf(x)\in L^1(\mathbf{R}^n)$,则 $\hat{f}(\xi)$ 对 $\xi_j$ 可求导,而且

$$\widehat{(x_jf)}(\xi)=\mathrm{i}\frac{\partial}{\partial\xi_j}\hat{f}(\xi). \tag{22}$$

(ii)若 $g(x)\in L^1(\mathbf{R}^n)$ 而它对某一自变量 $x_j$ 的不定积分 $f(x)$ 为 $L^1(\mathbf{R}^n)$ 函数,则 $f(x)=-\int_{x_j}^{\infty}g(x)\mathrm{d}x_j$,而且

$$\frac{1}{\mathrm{i}}\hat{g}(\xi)=\widehat{\left(\frac{1}{\mathrm{i}}\frac{\partial}{\partial\xi_j}f\right)}(\xi)=\xi_j\hat{f}(\xi). \tag{23}$$

**证** (i)记 $h_j$ 为第 $j$ 个分量为 $h$ 而其他分量为 0 的 $n$ 维向量.我们有

$$\hat{f}(\xi+h_j)=\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}x\xi}\cdot\mathrm{e}^{-\mathrm{i}x_jh_j}\mathrm{d}x,$$

所以

$$\frac{1}{h_j}[\hat{f}(\xi+h_j)-\hat{f}(\xi)]=\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}x\xi}\cdot\frac{\mathrm{e}^{-\mathrm{i}x_jh_j}-1}{h_j}\mathrm{d}x.$$

因为

$$|\mathrm{e}^{-\mathrm{i}x_jh_j}-1|=2\left|\sin\frac{x_jh_j}{2}\right|\leqslant|x_j||h_j|,$$

所以积分 $\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}x\xi}\cdot\frac{\mathrm{e}^{-\mathrm{i}x_jh_j}-1}{h_j}\mathrm{d}x$ 之被积函数绝对值不大于 $L^1(\mathbf{R}^n)$ 函数 $|xf(x)|$,于是我们可以应用勒贝格控制收敛定理在积分号下求极限而(22)得证.

(ii) 由假设 $f(x)=\int^{x_j} g(x)\mathrm{d}x_j$，现在我们利用 $f(x)\in L^1$ 这一条件来确定积分下限. 对任意 $A$ 与 $a$ 有

$$f(A)-f(a)=\int_a^A g(x)\mathrm{d}x_j.$$

因为假设了 $g(x)\in L^1(\mathbf{R}^n)$，所以当 $A\to\infty$ 时 $f(A)$ 有极限 $M$：$f(A)\to M$，同理当 $a\to-\infty$ 时 $f(a)\to m$. 又因为假设了 $f(x)\in L^1(\mathbf{R}^n)$，所以必有 $M=0$，$m=0$，而得

$$f(x)=-\int_{x_j}^{\infty} g(x)\mathrm{d}x_j.$$

对于勒贝格积分，我们知道不定积分在几乎处处意义下是原函数，所以 $g(x)=f'_{x_j}(x)$ 几乎处处成立. 分部积分公式

$$\int_a^A uv\mathrm{d}x=uV\Big|_a^A-\int_a^A Vu'\mathrm{d}x\quad (V\text{ 为 }v\text{ 之不定积分})$$

在勒贝格积分理论中当 $v$ 与 $u'$ 均为可积时是成立的，现在令 $v=g$，从而 $V=-\int_{x_j}^{+\infty} g(x)\mathrm{d}x_j=f(x)$ 以及 $u=\mathrm{e}^{-\mathrm{i}\xi x}$，我们有

$$\begin{aligned}\int_a^A g(x)\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x_j &=\int_a^A f'_{x_j}(x)\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x_j\\&=f(x)\mathrm{e}^{-\mathrm{i}x\xi}\Big|_a^A+\mathrm{i}\xi_j\int_a^A f(x)\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x.\end{aligned}$$

令 $a\to-\infty$，$A\to+\infty$，前面我们已经证明了 $f(a)\to0$，$f(A)\to0$，所以立即得到

$$\hat{g}(\xi)=\left(\widehat{\frac{\partial}{\partial x_j}f}\right)(\xi)=\mathrm{i}\xi_j\hat{f}(\xi),$$

从而 (23) 成立.

**注 1**　在黎曼积分理论中，我们是在两个条件下证明了 $\int_a^x f(x)\mathrm{d}x$ 是 $f(x)$ 的原函数的：其一是 $f(x)$ 黎曼可积，其二是 $f(x)$ 有原函数存在 (即处处适合 $F'(x)=f(x)$ 的 $F(x)$) 存在. 在勒贝格积分中我们也不能说只要 $f(x)\in L^1$，则 $\int_a^x f(x)\mathrm{d}x$ 一定是 $f(x)$ 的原函数，而只能得到 $\int_a^x f(x)\mathrm{d}x$ 在几乎处处意义下是 $f(x)$ 的原函数. 就是说，即令用勒贝格积分理论，也不能完全回答有可变积分限的积分 (即不定积分) 与原函数这两个概念是否完全一致，因为这里涉及一些比较细微问题，这是在这个定理证明中留下的一个"陷阱".

**注 2**　在本定理前，我们说了一句话："通常，要求一个函数 $\in L^1$ 可以归结为它在无穷远处衰减速度的限制." 实际上，事情当然不这么简单. 例如我们知道若 $\mathbf{R}$ 上的连续函数 $f(x)$ 在 $|x|\to\infty$ 时有 $f(x)=O\left(\frac{1}{x^\lambda}\right)$，$\lambda>1$，则 $f(x)$ 在 $\mathbf{R}$ 上绝对可积. 人们常常不自觉地以为其逆也是真的. 完全不然，甚至 $f(x)$ 当 $|x|\to\infty$ 时没有极限，甚至也不有界，仍有 $f(x)$ 在 $\mathbf{R}$ 上绝对可积. 这都可以找到例子. 这个定理的证明中，我们是在假设 $f(x)\in L^1$ 同时又是某 $L^1$ 函数 $g(x)$ 的不

定积分条件下才得知 $\lim\limits_{A\to\infty} f(A)=0$，$\lim\limits_{a\to-\infty} f(a)=0$ 的. 顺便说一下，从这里还得知，若 $g(x)\in L^1$，则 $\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x=0$ 时才能得到 $\int_{-\infty}^{x} g(x)\,\mathrm{d}x\in L^1$. 在这方面还有一个很有趣的结果(我们只对 $\mathbf{R}^1$ 的情况来陈述它)：若 $f(x)\in L^1$，$f^{(r)}(x)\in L^1$，$r>1$，则 $f^{(1)}(x),\cdots,f^{(r-1)}(x)$都$\in L^1$. 但是证明并不容易.

以上我们看到一个基本思想很清楚，而物理上又很重要的结论：求导与乘以自变量是对偶的，而且通过傅里叶变换来实现. 但是如果在 $L^1$ 框架下来讨论，会引起多少麻烦事. 因为广义函数框架在这方面显得更直观，物理学家和工程师很可能会乐于接受，以为这样就不需任何条件了. 但是情况并非如此，见下一节. 而且这丝毫不能说明 $L^1$ 和 $L^2$ 框架下的傅里叶变换理论只是烦琐哲学，恰好相反，它是内容十分丰富的.

最后我们再讲几个与自变量的线性变换有关的性质. 首先，也是最常用的是"反射"，即将 $x$ 变为 $-x$，或 $\xi$ 变为 $-\xi$，我们用一个专门记号来表示它，我们定义

$$\check{f}(x)=f(-x),\quad \check{\varphi}(\xi)=\varphi(-\xi).$$

首先讲一下如果 $\hat{f}(\xi)$ 仍在 $L^1(\mathbf{R}^n)$ 中，对它再作一次傅里叶变换又如何？事实上

$$\begin{aligned}\hat{\hat{f}}(x)&=\int_{\mathbf{R}^n}\hat{f}(\xi)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}\xi\\&=(2\pi)^n\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\hat{f}(\xi)\mathrm{e}^{\mathrm{i}\xi(-x)}\mathrm{d}\xi=(2\pi)^n f(-x)=(2\pi)^n\check{f}.\end{aligned}$$

注意这里自变量记号的变化，所以至少从形式上说，有

$$\hat{\hat{f}}(x)=(2\pi)^n\check{f}(x).\tag{24}$$

另一个重要的变换是平移算子. 我们定义平移算子 $\tau_h\,(h\in\mathbf{R}^n)$ 为

$$(\tau_h f)(x)=f(x-h).$$

很容易看到，若 $f\in L^1(\mathbf{R}^n)$，显然 $\tau_h f\in L^1(\mathbf{R}^n)$，易见

$$\begin{aligned}\widehat{(\tau_h f)}(\xi)&=\int_{\mathbf{R}^n}f(x-h)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x\\&=\mathrm{e}^{-\mathrm{i}\xi h}\int_{\mathbf{R}^n}f(x-h)\mathrm{e}^{-\mathrm{i}\xi(x-h)}\mathrm{d}x\\&=\mathrm{e}^{-\mathrm{i}\xi h}\hat{f}(\xi).\end{aligned}\tag{25}$$

我们知道 $\mathrm{e}^{-\mathrm{i}\xi h}$ 的 $\xi h$ 是相位，所以自变量的平移造成傅里叶变换相位变化. 再看若对 $f(x)$添加一个因子 $\mathrm{e}^{\mathrm{i}xh}$，则当 $f\in L^1(\mathbf{R}^n)$时，$\mathrm{e}^{\mathrm{i}xh}f(x)\in L^1(\mathbf{R}^n)$，容易看到

$$\begin{aligned}\widehat{(\mathrm{e}^{\mathrm{i}xh}f)}(\xi)&=\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}x(\xi-h)}\mathrm{d}\xi=\hat{f}(\xi-h)\\&=\tau_h\hat{f}(\xi).\end{aligned}\tag{26}$$

**2. $L^2(\mathbf{R}^n)$中的傅里叶变换**　首先遇到的问题是 $L^2(\mathbf{R}^n)$函数不一定可积，因此(6)式中的积分目前无定义. 但是 $L^2(\mathbf{R}^n)$是一个"很好"的空间：它有内积以及由内积导出的范数(因为现

在讨论的函数是取复值的,如 $e^{-i\xi x}$,所以我们应该讨论复的内积,即埃尔米特内积),它有正交性,因此可以定义其中的o.n.系.当我们考虑有限区间例如$[-\pi,\pi]$上的 $L^2$ 空间时,还可以找出可数的完全o.n.系如$\{e^{ikx}\}$,$k$ 为整数,而任一 $L^2[(-\pi,\pi)]$函数都可以按它展为傅里叶级数,且在 $L^2[(-\pi,\pi)]$中收敛. $L^2(\mathbf{R}^n)$空间又是完备的.这样一些性质使它在数学的许多分支和物理学中十分有用.它是所谓希尔伯特空间最常见的代表.因此,我们将着重从希尔伯特空间的角度来研究 $L^2(\mathbf{R}^n)$中的傅里叶变换.我们先从其定义开始,而暂时认为(6)只是一个形式的记法.

我们的作法是先取 $L^2(\mathbf{R}^n)$的一个稠密的子空间,使对于其中之元可以定义傅里叶变换,再应用 $L^2(\mathbf{R}^n)$的完备性把这个定义推广到整个 $L^2(\mathbf{R}^n)$.现在我们选取 $L^2(\mathbf{R}^n)$中具有紧支集$[-A,A]$的函数(为方便起见,我们只讨论一个自变量的情况,但是其结果易见可以适用于$\mathbf{R}^n$),并认为这些函数仍定义在 $\mathbf{R}$ 上不过在$[-A,A]$之外恒为0.记此空间为 $L^2(\mathbf{R}_A)$,很明显 $L^2(\mathbf{R}_A)\subset L^1(\mathbf{R}^n)$,所以可以在其上定义傅里叶变换.我们可以作一个相似变换

$$z=\frac{A}{\pi}y=\lambda y,\quad \lambda=\frac{A}{\pi},$$

使$[-A,A]$中的 $z$ 与$[-\pi,\pi]$中的 $y$ 相对应,而 $f(\lambda y)$之傅里叶变换

$$\begin{aligned}F[(\lambda\cdot)](\xi)&=\int_{-\infty}^{\infty}f(\lambda y)e^{-i\xi y}dy=\frac{1}{\lambda}\int_{-\infty}^{\infty}f(z)e^{iz\xi/\lambda}dz\\&=\frac{1}{\lambda}\hat{f}\left(\frac{\xi}{\lambda}\right).\end{aligned}$$

这当然不会影响下面的讨论.我们要注意现在我们是在 $L^2([-\pi,\pi])$中讨论问题.在$L^2[-\pi,\pi]$中有o.n.系$\left\{\frac{1}{\sqrt{2\pi}}e^{ikx}\right\}$,$k=0,\pm1,\pm2,\cdots$.在第四章就指出,这个o.n.系是完全的,但未证明.下面我们将把证明补出来,现在则承认这个结果.于是任意的 $f(x)\in L^2([-\pi,\pi])$均可展开为 $L^2$ 意义下收敛的傅里叶级数.于是任意的 $f(x)\in L^2([-\pi,\pi])$均可展开为在 $L^2$ 意义下收敛的傅里叶级数:

$$f(x)=\sum_{k=-\infty}^{+\infty}a_k\frac{1}{\sqrt{2\pi}}e^{ikx},$$

这里

$$a_k=\int_{-\pi}^{\pi}f(x)\cdot\frac{1}{\sqrt{2\pi}}e^{-ikx}dx=\frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}f(x)e^{-ikx}dx=\frac{C_k}{\sqrt{2\pi}},$$

而且

$$C_k=\int_{-\pi}^{\pi}f(x)e^{-ikx}dx$$

正是通常说的傅里叶系数.而$\{e^{ikx}\}$之完全性给出了以下的帕塞瓦尔关系式

$$\|f\|_{L^2}^2=\sum_{k=-\infty}^{+\infty}|a_k|^2=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\left|\int_{-\pi}^{\pi}f(x)e^{-ikx}dx\right|^2,$$

但右方的积分正是$|\hat{f}(k)|^2$.所以我们有

$$\|f\|_{L^2}^2=\frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}|\hat{f}(k)|^2.\tag{27}$$

如果我们用 $e^{-i\lambda x}f(x)$，$0\leqslant\lambda\leqslant 1$，代替 $f(x)$则其傅里叶变换成为$\hat{f}(k+\lambda)$. 但是 $\|e^{-i\lambda x}f(x)\|_{L^2}^2=\|f(x)\|_{L^2}^2$. 因此又有

$$\|f\|_{L^2}^2=\frac{1}{2\pi}\sum_{k=-\infty}^{+\infty}|\hat{f}(k+\lambda)|^2. \tag{28}$$

这是一个极重要的关系式. 双方对 $\lambda$ 由 0 到 1 积分，立即知道，对于支集在$[-\pi,\pi]$中的 $L^2$ 函数，有

$$\|f\|_{L^2}^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{f}(\xi)|^2\mathrm{d}\xi=\frac{1}{2\pi}\|\hat{f}\|_{L^2}^2. \tag{29}$$

利用它，首先我们就可以定义 $L^2(\mathbf{R}^n)$函数的傅里叶变换. 仍令 $n=1$，作

$$f_N(x)=\begin{cases}f(x), & |x|\leqslant N,\\ 0, & |x|>N,\end{cases}$$

考虑 $\varphi(x)=f_N(x)-f_{N'}(x)$，不妨设 $N'<N$，于是 $f_N(x)-f_{N'}(x)$之支集在$[-N,N]$中. 如上所述，作相似变换 $x=\frac{N}{\pi}y=\lambda y$，$\varphi(\lambda y)$作为 $y$ 的函数，支集在$[-\pi,\pi]$中，其傅里叶变换则成了$\frac{1}{\lambda}\hat{\varphi}\left(\frac{\xi}{\lambda}\right)$.

$$\begin{aligned}\|f_N(x)-f_{N'}(x)\|_{L^2}^2&=\int_{-N}^{N}|\varphi(x)|^2\mathrm{d}x=\lambda\int_{-\pi}^{\pi}|\varphi(\lambda y)|^2\mathrm{d}y\\&=\lambda\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{1}{\lambda}\hat{\varphi}\left(\frac{\xi}{\lambda}\right)\right|^2\mathrm{d}\xi=\frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{\varphi}(\xi)|^2\mathrm{d}\xi\\&=\frac{1}{2\pi}\|\hat{f}_N(\xi)-\hat{f}_{N'}(\xi)\|_{L^2}^2.\end{aligned}$$

但是$\{f_N(x)\}$是 $L^2(\mathbf{R}_x)$中的柯西序列，而且其 $L^2$ 极限就是 $f(x)\in L^2$. 所以$\{\hat{f}_N(\xi)\}$也是 $L^2(\mathbf{R}_\xi)$中的柯西序列. 我们就以它的极限作为 $f(x)$之傅里叶变换，并仍记为$\hat{f}(\xi)$. 于是我们得到了

**定义 2** 设 $f(x)\in L^2(\mathbf{R}_x^n)$，则

$$\hat{f}_N(\xi)=\int_{-N}^{N}f(x)e^{-ix\xi}\mathrm{d}x$$

在 $L^2(\mathbf{R}_\xi)$中有极限，我们称之为 $f(x)$之傅里叶变换，并仍记以$\hat{f}(\xi)$，显然$\hat{f}(\xi)\in L^2(\mathbf{R}_\xi)$：

$$\hat{f}(\xi)=\underset{N\to\infty}{\mathrm{l.i.m.}}\int_{\mathbf{R}^n}f_N(x)e^{-i\xi x}\mathrm{d}x.$$

l.i.m.是 limit in mean(平均收敛)的缩写.

用这样的方法需要补足上一章§4 中留下的一个漏洞，即 $\left\{\frac{1}{\sqrt{2\pi e^{ikx}}}\right\}k=0,\pm1,\pm2,\cdots$ 是 $L^2([-\pi,\pi])$中的一个完全 o.n.系，证明如下.取$[-A,A]\subset[-\pi,\pi]$，这里 $\pi-A>\delta>0$，而且作一个新的 $L^2([-\pi,\pi])$函数：

$$f_A(x)=\begin{cases}f(x), & x\in[-A,A],\\ 0, & A\leqslant|x|\leqslant\pi.\end{cases}$$

然后用磨光核 $j_\eta(x-y)=\dfrac{1}{C\eta}j_1\left(\dfrac{x-y}{\eta}\right)$对 $f_A(x)$作磨光:

$$g_\eta(x)=(J_\eta f_A)(x)=\frac{1}{C\eta}\int_{-\infty}^{\infty}j_1\left(\frac{x-y}{\eta}\right)f_A(y)\mathrm{d}y.$$

这里 $\eta<\delta$,而 $C$ 的选取使

$$\frac{1}{C}\int_{-\infty}^{\infty}j_1(x)\mathrm{d}x=1.$$

于是 $g_\eta(x)\in C^\infty$,而且 $g_\eta(x)$之支集在$[-A-\eta,A+\eta]$之内,所以 $g_\eta(x)$是$[-\pi,\pi]$中的 $C_0^\infty$ 函数,且由前面的说明,$g_\eta(x)$在 $L^2$ 中收敛于 $f(x)$,故对任一 $\varepsilon>0$,必有 $\eta$,使

$$\| g_\eta-f\|_{L^2}<\frac{\varepsilon}{2}.$$

现在我们固定 $\eta$.

因为 $g_\eta(x)\in C_0^\infty$,所以 $g_\eta(x)$的傅里叶级数一致收敛于 $g_\eta(x)$. 于是必存在一个 $N_0$,当 $N>N_0$ 时

$$\left\| g_\eta(x)-\sum_{k=-N}^{N}\frac{q_k}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}\right\|^2=\int_{-\pi}^{\pi}\left|g_\eta(x)-\sum_{k=-N}^{N}\frac{q_k}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}\right|^2\mathrm{d}x<\left(\frac{\varepsilon}{2}\right)^2.$$

这里 $q_k$ 是 $g_\eta$ 对 o.n.系$\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}$的傅里叶系数.

但由贝塞尔不等式,若 $C_k=\left(f,\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}\right)$时,$\left\| f-\sum\limits_{k=-N}^{N}\dfrac{C_k}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}\right\|\leqslant\left\| f-\sum\limits_{k=-N}^{N}\dfrac{q_k}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}\right\|\leqslant \| f-g\|+\left\| g-\sum\limits_{k=-N}^{N}\dfrac{q_k}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}\right\|<\varepsilon$. 由 $\varepsilon$ 之任意性,知 $\sum\limits_{-\infty}^{\infty}\dfrac{C_k}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}kx}$在 $L^2$ 中收敛于 $f(x)$,这就是说$\left\{\dfrac{\mathrm{e}^{-\mathrm{i}kx}}{\sqrt{2\pi}}\right\}$是一个完全的 o.n.系.

(28)式的另一个应用是以下极重要的

**定理 5(普兰舍利(Plancherel)定理)**　设 $f(x)\in L^2(\mathbf{R}_x)$,则$\hat{f}(\xi)\in L^2(\mathbf{R}_\xi)$,而且二者范数"相等".

$$\| f\|_{L^2(\mathbf{R}_x)}^2=\frac{1}{2\pi}\|\hat{f}\|_{L^2(\mathbf{R}_\xi)}^2. \tag{30}$$

**证**　证明极简单:对 $f_N(x)$应用(29)式,如上所述会有

$$\| f_N\|_{L^2}^2=\frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{f}_N(\xi)|^2\mathrm{d}\xi.$$

这就是说,对紧支集的 $L^2$ 函数,普兰舍利定理是成立的,令 $N$

这个因子就消失了,或者在 $\mathbf{R}_\xi$ 空间中用$\frac{1}{2\pi}\mathrm{d}\xi$ 作为其测度亦可. 这只是一个习惯问题. 但我们宁愿用语上“马虎”一些. 例如下面要讲的“等距”变换等等,其实也有时差一个因子$\frac{1}{2\pi}$,读者在计算时小心一点就行了. 又如上文中我们作积分时有时是把 $x$ 作为一维变量,而积分限又写为 $\mathbf{R}^n$,都是如此. 总之,这是一个习惯问题.

普兰舍利定理意义重大,它告诉我们傅里叶变换(我们暂时改用一个记号 $F$:$(Ff)(\xi)=\hat{f}(\xi)$)是由 $L^2(\mathbf{R}_x)$ 到 $L^2(\mathbf{R}_\xi)$ 的线性变换,而且使得范数不变(其实差一个因子$\frac{1}{2\pi}$). 不仅如此,它还保持内积不变(注意现在用复值系数,所以内积指的是埃尔米特内积,其记号也采用埃尔米特配对):

$$(f,g)=\frac{1}{2\pi}(\hat{f},\hat{g}). \tag{31}$$

由范数不变推导出内积不变的办法是标准的,称为极化关系式(polar relations)如下:令 $f,g\in L^2(\mathbf{R}^n)$则

$$\|f+g\|^2=\|f\|^2+\|g\|^2+(f,g)+(g,f),$$
$$\|f+\mathrm{i}g\|^2=\|f\|^2+\|g\|^2-\mathrm{i}(f,g)+\mathrm{i}(g,f).$$

用 i 乘后式再加到前式,有

$$2(f,g)=[\|f+g\|^2-\|f\|^2-\|g\|^2]+\mathrm{i}[\|f+\mathrm{i}g\|^2-\|f\|^2-\|g\|^2].$$

把 $f,g$ 分别变成$\hat{f},\hat{g}$ 并在双方都加一个因子$\frac{1}{2\pi}$,再用(30)即得

$$(f,g)=\frac{1}{2\pi}(\hat{f},\hat{g}).$$

一个希尔伯特空间(如 $L^2(\mathbf{R}^n)$)到其自身的线性算子 $T$,如果是此空间的线性同构,且能保持埃尔米特内积不变(在傅里叶变换的情况下要加中一上因子$\frac{1}{(2\pi)^n}$,如(31)式):

$$(Tf,Tg)=(f,g) \tag{32}$$

就称为一个酉算子(unitary operator),所以,普兰舍利定理启发我们,$L^2(\mathbf{R}^n)$上的傅里叶变换是一个酉算子. 其实在线性代数中我们已经看过了什么是酉矩阵:例如,如果 $f,g$ 是 $n$ 维复线性空间中的 $n$ 维向量,$T$ 看成 $n\times n$ 复数元矩阵,则由伴随矩阵 $T^*=(t_{ij})^*=(\overline{t_{ji}})$之定义,(32)告诉我们

$$T^*T=I,TT^*=I,$$

亦即 $T^*=T^{-1}$. 所以在有限维情况下,酉算子一定可逆. 现在转到 $L^2(\mathbf{R}^n)$,它是一个(可数)无限维希尔伯特空间,一个复希尔伯特空间中的酉算子 $T$ 即一个保持范数不变的线性同构. 那么,傅里叶变换作为复希尔伯特空间 $L^2(\mathbf{R}^n)$到其自身的线性变换,是否是一个酉算子呢?当然它是一个线性算子,而且普兰舍利定理又保证了它保持范数不变. 余下的只有“同构”待考. 同构包含了两个意思,一是单射. 即只有 $0\in L^2(\mathbf{R}_x^n)$才被映为 $0\in L^2(\mathbf{R}_\xi^n)$. 这样就不会有两个 $L^2(\mathbf{R}_x^n)$函数 $f_1(x)$和 $f_2(x)$被映为同一个$\hat{f}(\xi)$. 因为如果 $f_1(x)$和 $f_2(x)$有相同的$\hat{f}_1(\xi)=$

$\hat{f}_2(\xi)$，则 $\varphi(x)=f_1(x)-f_2(x)$之傅里叶变换$\hat{\varphi}(\xi)=\hat{f}_1(\xi)-\hat{f}_2(\xi)=0$. 但由定理 5，这时

$$\|\varphi\|^2_{L^2(\mathbf{R}^n_x)}=\frac{1}{(2\pi)^2}\|\hat{\varphi}\|^2_{L^2(\mathbf{R}^n_\xi)}=0,$$

所以 $\varphi=f_1(x)-f_2(x)=0$. 其次要证明它是满射. 即任一 $L^2(\mathbf{R}^n_\xi)$函数 $\varphi(\xi)$必是某一 $L^2(\mathbf{R}^n_x)$函数 $f(x)$之傅里叶变换. 我们来证明

$$f(x)=\frac{1}{(2\pi)^n}\int e^{i\xi x}\varphi(\xi)d\xi$$

恰好合于所求. 但这正是反演公式，我们现在对 $n=1$ 的情况来"证明"它. (31)左方是

$$\int f(x)\overline{g(x)}dx,$$

而右方为

$$\begin{aligned}&\frac{1}{(2\pi)^n}\int d\xi\left[\int f(y)e^{-i\xi y}dy\cdot\overline{\int g(x)e^{-i\xi x}dx}\right]\\&=\frac{1}{(2\pi)^n}\int d\xi\iint dyf(y)\overline{g(x)}e^{-i\xi(y-x)}dx\\&=\frac{1}{(2\pi)^n}\int\left[\int e^{i\xi x}d\xi\int f(y)e^{-i\xi y}dy\cdot\overline{g(x)}\right]dx.\end{aligned}$$

比较双方，利用 $g(x)$之任意性即有

$$f(x)=\frac{1}{(2\pi)^n}\int e^{i\xi x}d\xi\int f(y)e^{-i\xi y}dy.$$

这就证明了(6)确实是(11)的逆变换公式.

这个"证明"是不合法的，因为交换积分次序还需论证，我们将在下一节讨论它.

以上我们只是对 $L^2(\mathbf{R}^n)$中的傅里叶变换之定义及其基本性质(它是酉算子，并有普兰舍利定理)作了详细讨论. 但还留下了一个漏洞，即逆傅里叶变换公式未严格证明. 傅里叶变换的另一些初等的性质，对于 $L^2$ 和 $L^1$ 是一样的，有待证明，但其中有一些我们仍留待下一节用广义函数来处理.

首先来看卷积，若 $f(x), g(x)\in L^2(\mathbf{R}^n)$，则若视 $x$ 为一参数，$f(t)g(t-x)$作为 $t$ 的函数是$L^1(\mathbf{R}^n)$函数. 因此$\int_{-\infty}^{\infty}f(t)g(t-x)dt$ 是有意义的，它定义了一个 $x$ 的函数$h(x)$，但是关于 $h(x)$的性质，除了易证它是有界的：

$$\begin{aligned}|h(x)|&\leqslant\int_{-\infty}^{+\infty}|f(t)||g(t-x)|dt\leqslant\left(\int_{-\infty}^{\infty}|f(t)|^2dt\right)^{\frac{1}{2}}\cdot\left(\int_{-\infty}^{+\infty}|g(t-x)|^2dt\right)^{\frac{1}{2}}\\&=\|f\|_{L^2}\cdot\|g\|_{L^2}\end{aligned}$$

以外，我们实在所知不多，所以我们不考虑如何定义两个 $L^2(\mathbf{R}^n)$函数的卷积.

其次看一下求导问题. 有一些很细致的定理，例如(我们只讲一维情况，$n$ 维的不讲了)

**定理 6** 设 $f(x)\in L^2(\mathbf{R})$是绝对连续的，而且 $f'(x)\in L^2(\mathbf{R})$，则 $\hat{f'}(\xi)=i\xi\hat{f}(\xi)$，它自然仍在 $L^2(\mathbf{R})$中.

**定理 7** 若 $f(x), g(x)\in L^2(\mathbf{R})$，而且 $\hat{g}(\xi)=\xi\hat{f}(\xi)$，则 $g(x)$为局部绝对连续(即在每一

个有限区间上为绝对连续)，且

$$g(x)=f'(x);$$

$$f(x)=-\int_x^{+\infty}g(y)\mathrm{d}y.$$

这些定理我们都未证明，留待用广义函数来讨论.

**3. 傅里叶变换与量子力学**　现在我们从量子力学角度来看一下普兰舍利定理. 首先，我们想关于量子力学稍微说几句. 19 世纪末的物理学的许多新的实验事实与经典物理学的基本概念相矛盾，使得普朗克在 1900 年提出能量子假说，即能量不是连续的，而是由一些极小的离散单位——即能量子——构成的. 爱因斯坦在研究光电现象时进一步发展了这个思想，提出光量子——即光子——不但能量 $E$ 是分立的，而且动量 $p$ 也是分立的，而且能量与动量与光作为一种波动的宏观的物理量频率 $\nu$ 及波长 $\lambda$ 有关：

$$E=h\nu,\ p=h/\lambda,\ h=6.626\times10^{-34}\ \mathrm{J\cdot s}.$$

这一些都经过了实验的验证. 这里的 $h$ 称为普朗克常量，是量子现象中最重要的普适常量. 另一个重大问题是原子的构造. 人们曾经认可的原子行星模型——中间有核，外层有电子，而且服从开普勒的行星运动规律——是 1911 年卢瑟福(E. Rutherford)提出的. 然而按经典物理的定律，这样的原子是不稳定的，其寿命的数量级只有 $10^{-9}$秒. 原子辐射的光谱又有明显的明线与暗线，这也与经典物理不相容. 波尔(N. Bohr)对此有解释：外层电子的轨道并不是任意的，而只能处于一系列固定的能级上. 当电子在一个轨道上运行时，一切都很正常，也不会辐射能量. 但若它从一个轨道跃迁到其它能级的轨道，则将吸收或辐射出一定的能量，等于两轨道能级之差. 这时就产生光谱中的明线(发射能量)或暗线(吸收能量). 但是当时物理学家的方法论还是承认经典物理的定律，不过要在某些特定情况下加以修正. 所以波尔称这种理论为“旧量子理论”.

到了 20 世纪 20 年代，人们开始认识到微观世界有自己的物理规律. 所谓微观世界，人们当然都知道，例如亚原子世界当然是微观世界. 但是我们更准确地可以说，量子力学处理的问题即普朗克常量 $h$ 不能忽略的问题. 人们把微观世界的规律归结为若干要点，而在 20 世纪 30 年代由冯•诺依曼(J. Von Neumann)归结为几个公设作为从数学上处理量子力学的基础. 与牛顿时代提出三大定律(牛顿也是把它们当作数学公理看待的)不同，量子力学的公设更具抽象性与某项具体的物理实验距离更远. 我们下面就提出几条“公设”而不加“证明”. 我们没有用几个基本命题的形式来陈述这些公设，而是就我们所讨论的问题作了一些解释.

1. 量子力学系统的状态可以用状态空间中的波函数 $\Psi(x,t)$来刻画. $\Psi$ 取复值而与 $c\Psi$ ($c\neq0$ 是一个复数)表示同一个态. 在每一时刻 $t$，$\Psi\in L^2(\mathbf{R}^n)$，所以 $\Psi$ 可以规范化，即适合 $\|\Psi\|_{L^2}=1$. 但即令如此，$\Psi$ 仍不能唯一决定：$\Psi$ 与 $\mathrm{e}^{\mathrm{i}\theta}\Psi$ ($\theta$ 为实数)仍表示同一个态. 态空间是一个复希尔伯特空间，具体说来即是 $L^2(\mathbf{R}^n)$.

2. 我们只研究可以观测到的物理量. 所以量子力学讲的物理量都称为可观测量(observables). 例如粒子的位置、动量、能量等都是可观测量. 但是时间 $t$ 则作为一个参数处理而不算是可观测量.

微观世界的对象如此之小，要观测它们就不能忽略观测仪器对于对象的作用. 观测的结果一般的是一个随机变量，就是可以取许多值的量，但是它取某个范围中的值的概率是一定的. 例如，若波函数为 $\Psi(x,t)$，则粒子位于 $x$ 与 $x+\mathrm{d}x$ 之间的概率为 $|\Psi(x,t)|^2\mathrm{d}x$.

以上是讲的观测结果．至于可观测量本身，在经典物理中是用实数来表示的，但量子力学中的可观测量则用态空间 $L^2(\mathbf{R}^n)$上的自伴算子来表示．例如粒子的位置在经典物理中用坐标 $x$ 表示，在量子物理中则是乘以 $x$：它作用在波函数 $\varphi(x)\in L^2(\mathbf{R}^n)$上给出 $x\cdot\varphi(x)$．所以 $x$ 很像线性代数中的记号 $xI$，但我们仍习惯地用 $x$ 来表示这个算子．动量算子则为$\frac{1}{\mathrm{i}}\partial_x$，我们常用 $p$ 或 $\xi$ 表示．这里就出现一个问题，以算子 $x$ 为例，$\varphi(x)\in L^2(\mathbf{R}^n)$，但 $x\varphi(x)$不一定仍在 $L^2(\mathbf{R}^n)$中．因此，凡讲到算子，其定义域不一定是整个态空间 $L^2(\mathbf{R}^n)$，从数学上讲，至少我们需要这些算子的定义域在 $L^2(\mathbf{R}^n)$中稠密．至于动量算子更有问题了：$p\varphi(x)=\frac{1}{\mathrm{i}}\partial_x\varphi(x)$，而 $\varphi(x)$不一定可微甚至不是广义可微，所以 $p$ 的定义域更不可能是整个$L^2(\mathbf{R}^n)$．这一点我们在下面讲测不准原理时就会遇到．在那里，我们不去追问例如为什么不要求 $\varphi(x)$适合 $x\varphi(x)\in L^2(\mathbf{R}^n)$？我们既把 $\varphi(x)$放在算子 $x$ 的作用之下，就是承认了 $\varphi(x)$在 $x$ 的定义域中，亦即承认 $x\varphi(x)\in L^2(\mathbf{R}^n)$．

所谓自伴算子 $A:L^2(\mathbf{R}^n)\to L^2(\mathbf{R}^n)$，简单说就是对于定义域中的 $\varphi(x)$与 $\psi(x)$，适合关系式

$$(A\varphi,\psi)=(\varphi,A\psi),$$

$$\int_{\mathbf{R}^n}(A\varphi)(x)\overline{\psi(x)}\,\mathrm{d}x=\int_{\mathbf{R}^n}\varphi(x)\overline{(A\psi)(x)}\,\mathrm{d}x$$

的算子，由此易见$(A\varphi,\varphi)$必定取实值．

3．前面说了观测的结果是一个随机变量，例如我们要在波函数 $\varphi(x)$所表示的状态下问粒子位置 $x$(现在的 $x$ 可不代表算子了)，则因 $x$ 的值在$x$ 与$x+\mathrm{d}x$ 之间的概率为$|\varphi(x)|^2\mathrm{d}x$，所以考虑一切可能的测量结果，应该有一个平均值．这在数学上称为数学期望值，它应该是$\int x|\varphi(x)|^2\mathrm{d}x=(x\varphi,\varphi)$(最后一项中的 $x$ 是代表算子了，它作用在波函数 $\varphi(x)$上的结果是$x\cdot\varphi(x)$)．同样，在状态 $\varphi$ 之下，可观测量 $A$ 的平均值即数学期望值$(A\varphi,\varphi)$记作 $a_\varphi$．由上面的自伴算子的性质，$a_\varphi=(A\varphi,\varphi)$取实数值．在物理上这是唯一可接受的：观测值必须是实数．

既然观测的结果是随机变量，而又有一个平均值．那么，观测结果与平均值的偏离程度或者说，观测结果分散的程度就应该用方差(或称离差，其平方根称为标准差)来刻画．在上一章讲高斯积分时就讲到，方差应该是每一次观测结果与平均值(即数学期望值)之差的平方之平均值．用在我们的情况，则在状态 $\varphi$ 之下，可观测量 $A$ 之方差$\delta_\varphi A$ 应该为

$$\begin{aligned}\delta_\varphi A&=(A-a_\varphi)^2\text{ 之平均值}\\&=((A-a_\varphi)^2\varphi,\varphi)=((A-a_\varphi)\varphi,(A-a_\varphi)\varphi)\\&=\|(A-a_\varphi)\varphi\|^2=\|(A-a_\varphi I)\varphi\|^2.\end{aligned}$$

上式中的 $a_\varphi$ 本来是一个实数，但这里要把它看作一个算子，正如在线性代数中我们若写 $A-\lambda$ 就必须理解为$A-\lambda$

4. $L^2(\mathbf{R}^n)$上的酉算子 $U$ 即适合 $U^* U = I$ 的线性算子. 这时因为 $U^{-1} = U^*$ 作用在 $L^2(\mathbf{R}^n)$上,所以 $U$ 的值域是$L^2(\mathbf{R}^n)$全空间,同时 $U^{-1}$的存在,又说明 $U$ 是 $\mathbf{1-1}$ 的,是一个线性同构. $U$ 算子又保持内积不变:

$$(U\varphi, U\varphi) = (U^* U\varphi, \varphi) = (\varphi, \varphi).$$

所以,$L^2(\mathbf{R}^n)$在经过 $U$ 的作用后,其线性构造与度量都不变,因此 $L^2(\mathbf{R}^n)$没有改变. 现在看傅里叶变换在 $L^2(\mathbf{R}^n)$上的作用. 普兰舍利定理已证明了傅里叶变换保持内积不变. 在证明了逆变换定理后(上面我们只是“形式地”推算了一下),又知道傅里叶变换保持 $L^2(\mathbf{R}^n)$的线性构造不变,而且是 $\mathbf{1-1}$ 的从而是一个酉算子. 这样,$\varphi(x)$是一个波函数而表示一个量子系统的一个状态,则$\hat{\varphi}(\xi)$ 也应该表示同一状态. 前者称为该状态的 $x$ 表象,后者称为其 $\xi$ 表象. 按波函数的概率解释,$|\hat{\varphi}(\xi)|^2$ 应该表示 $\xi$ 的观测值在$\xi$与$\xi + \mathrm{d}\xi$之间的概率密度——$|\hat{\varphi}(\xi)|^2\mathrm{d}\xi$ 表示概率. 现在要问,$\xi$ 在物理上表示什么? 为此我们来看动量算子$\frac{1}{\mathrm{i}}\partial_x$ 在傅里叶变换下表示什么. 至少,形式的计算告诉我们,通过分部积分

$$\int\limits_{\mathbf{R}^n} \frac{1}{\mathrm{i}}\partial_{x_j}\varphi(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x = \xi_j \int\limits_{\mathbf{R}^n} \varphi(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x = \xi_j\hat{\varphi}(\xi),$$

积分号外之项变为 0,前面我们已对 $L^1(\mathbf{R}^n)$证明了,若 $\varphi(x)$和$\frac{\partial}{\partial x_j}\varphi(x)$均为可积则当$|x_j|\to\infty$时,$\varphi(x)\to 0$. $L^2(\mathbf{R}^n)$的情况下则需另证,所以在这里我们得到的只能算是形式推导,即是说 $\xi$ 表象下的乘以$\xi_j$ 与$x$ 表象下的动量算子$\frac{1}{\mathrm{i}}\partial_{x_j}$是一样的. 所以 $\xi$ 表象又称动量表象.

5. 在经典力学中粒子的位置和动量(或简单地就说是速度)共同决定了该粒子的状态. 量子物理中却只用 $x$($x$ 表象)或 $\xi$($\xi$ 表象),这至少是因为,如果 $x$ 与$\xi$共同才决定粒子的状态,则必须要能同时观测 $x$ 与 $\xi$. 但是这是不可能的. 因为量子力学的基础公设之一就是两个可观测量要能同时观测的充分必要条件是它们所对应的自伴算子 $A$ 与$B$ 要可交换,即 $AB - BA$

这里我们引进一个在数学文献中常用的概念“交换子”(commutator):两个算子 的交换

所以，$\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x_j}$(动量的第 $j$ 个分量)与 $x_j$ (位置的第 $j$ 个分量)是不能同时观测的. 如果对这两个可观测量同时进行了观察，它必有不可任意减小的误差. 下面把这段话准确一些叙述. 在状态 $\varphi$ 下观测 $A$ 与 $B$ 的结果有平均值 $a_\varphi$ 和 $b_\varphi$，而观测结果是分散开的，我们用方差 $\delta_\varphi A = \|(A-a_\varphi)\varphi\|^2$ 与 $\delta_\varphi B = \|(B-b_\varphi)\varphi\|^2$ 来度量，我们还引用了标准差 $\Delta_\varphi A=\sqrt{\delta_\varphi A}$，$\Delta_\varphi B=\sqrt{\delta_\varphi B}$，于是我们有

**定理 8 Ⅰ**（**测不准原理Ⅰ**）　在状态 $\varphi$ 之下，我们有

$$\Delta_\varphi A\cdot\Delta_\varphi B\geqslant\frac{1}{2}|([A,B]\varphi,\varphi)|. \tag{33}$$

**证**　令 $A_1=A-a_\varphi I, B_1=B-b_\varphi I$，很容易验证 $[A,B]=[A_1,B_1]$ 因此

$$\begin{aligned}|([A,B]\varphi,\varphi)| &= |([A_1,B_1]\varphi,\varphi)| = |(A_1B_1\varphi,\varphi)-(B_1A_1\varphi,\varphi)|\\ &= |(B_1\varphi,A_1\varphi)-(A_1\varphi,B_1\varphi)| = |2\mathrm{Im}(A_1\varphi,B_1\varphi)|\\ &\leqslant 2\|A_1\varphi\|\cdot\|B_1\varphi\| = 2\|(A-a_\varphi)\varphi\|\cdot\|(B-b_\varphi)\varphi\|\\ &= 2\Delta_\varphi A\cdot\Delta_\varphi B.\end{aligned}$$

这样，两个标准差之积的下限依赖于状态. 但有一个情况可以改进这个情况，这就是可观测量 $A, B$ 为典则共轭的情况，就是 $[A,B]=\hbar I/\mathrm{i}$ 的情况. 例如动量算子 $p_j=\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial x_j}$ 与同标号的位置算子 $x_j$ 就是这样，这时有

**定理 8 Ⅱ**（**测不准原理Ⅱ**）　若 $A, B$ 为典则共轭的可观测量，则在任意状态 $\varphi(x)$ 下

$$\Delta_\varphi A\cdot\Delta_\varphi B\geqslant\frac{\hbar}{2}. \tag{34}$$

**证**　(33)的右方现在成为

$$\frac{1}{2}\left|\left(\frac{\hbar}{\mathrm{i}}\varphi,\varphi\right)\right|=\frac{\hbar}{2}\|\varphi\|=\frac{\hbar}{2}.$$

故(34)得证.

特别是动量与位置恰好是典则共轭的，故

$$\Delta p_j\cdot\Delta x_j\geqslant\frac{\hbar}{2}. \tag{35}$$

本来我们是在讨论 $L^2(\mathbf{R}^n)$ 中的傅里叶变换，但现在似乎转成了希尔伯特空间中自伴算子的性质. 但是至少在动量与位置的测不准关系上，确实只须用傅里叶变换的性质就够了. 其中的要点正是普兰舍利定理. 除此以外又需利用傅里叶变换的简单性质. 下面再就 $p_j$ 与 $x_j$ 证明一次测不准原理(35). 我们引入 $A_1=A-a_\varphi I$，这等于假设了 $a_\varphi=0$，用于位置算子 $x_jI$ 就相当于说位置的平均值 $x_j^0=0$. 其实，即令 $x_j$ 之平均值 $x_j^0\neq0$，我们可以作一个平移 $y_j=x_j-x_j^0$，这时，波函数 $\varphi(x)=\varphi(y_1+x_1^0,\cdots,y_n+x_n^0)=\varphi(y+x^0)$，于是其傅里叶变换成为

$$\int\varphi(y+x^0)\mathrm{e}^{-\mathrm{i}x\xi}\mathrm{d}x=\int\varphi(y+x^0)$$

$\delta_{\varphi}(\hbar\xi_j)=\frac{1}{2\pi}\hbar^2\int\xi_j^2|\mathrm{e}^{\mathrm{i}x^0\xi}\hat{\varphi}(\xi)|^2\mathrm{d}\xi=\frac{\hbar^2}{2\pi}\int\xi_j^2|\hat{\varphi}(\xi)|^2\mathrm{d}\xi=\frac{\hbar^2}{2\pi}\int|\partial_j\hat{\varphi}(\xi)|^2\mathrm{d}\xi$. 同样，若动量 $\hbar\xi_j$ 之平均值为 $\hbar\xi_j^0(j=1,2,\cdots,n)$，在 $\xi$ 空间中作平移 $\xi_j=\xi_j-\xi_j^0$，波函数将成为 $\hat{\varphi}(\xi+\xi^0)$，回到 $x$ 表象后将得到波函数 $\mathrm{e}^{-\mathrm{i}\xi^0x}\varphi(x)$，这同样不影响到概率密度 $|\mathrm{e}^{-\mathrm{i}\xi^0x}\varphi(x)|^2=|\varphi(x)|^2$，所以 $\delta_\varphi$ 也不变：

$$\delta_{\varphi}(x_j)=\int x_j^2|\varphi(x)|^2\mathrm{d}x.$$

现在我们需要证明的只有

$$\delta_{\varphi}(\hbar\xi_j)\cdot\delta_{\varphi}(x_j)\geqslant\frac{\hbar^2}{4}.$$

这里的 $\varphi\in L^2(\mathbf{R}^n)$，而且 $x\varphi(x)\in L^2(\mathbf{R}_x^n)$，$\frac{\partial}{\partial x_j}\varphi(x)\in L^2(\mathbf{R}_x^n)$. 和前面一样，我们只考虑一维 $(n=1)$ 的情况，而且只对支集在 $-N\leqslant x\leqslant N$ 中的 $\varphi(x)$ 证明上式，然后再取极限即可. 为此，考虑积分.

$$I(\lambda)=\int_{-\infty}^{+\infty}|\lambda x\varphi(x)+\varphi'(x)|^2\mathrm{d}x=\int_{-\infty}^{+\infty}[\lambda x\varphi(x)+\varphi'(x)]\overline{[\lambda x\varphi(x)+\varphi'(x)]}\mathrm{d}x.$$

这里 $\lambda$ 是实数，$I(\lambda)$ 显然 $\geqslant 0$，而且是 $\lambda$ 的二次三项式：

$$I(\lambda)=\lambda^2\int_{-\infty}^{+\infty}x^2|\varphi(x)|^2\mathrm{d}x+\lambda\int_{-\infty}^{\infty}x(\varphi\bar{\varphi}'+\bar{\varphi}\varphi')\mathrm{d}x+\int_{-\infty}^{\infty}|\varphi'(x)|^2\mathrm{d}x,$$

因此其判别式非正. 但是 $\varphi\bar{\varphi}'+\bar{\varphi}\varphi'=(\varphi\bar{\varphi})'$ 故

$$\int_{-\infty}^{+\infty}x(\varphi\bar{\varphi}'+\bar{\varphi}\varphi')\mathrm{d}x=x\varphi\bar{\varphi}\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}|\varphi(x)|^2\mathrm{d}x=-1.$$

注意，我们这里利用了 $\varphi(x)$ 支集在 $-N<x<N$ 中且波函数是规范化了的. 于是我们有

$$0\geqslant 1-4\int_{-\infty}^{\infty}x^2|\varphi(x)|^2\mathrm{d}x\cdot\int_{\infty}^{\infty}|\varphi'(x)|^2\mathrm{d}x$$

这里 $C$ 应保证 $\varphi(x)$ 规范化：$\|\varphi(x)\|=1$，故 $C=(\sqrt{\lambda}\pi)^{\frac{1}{2}}$. 所以，测不准原理只有当波函数 $\varphi(x)$ 为高斯函数时才有等号成立.

测不准原理既然如此深刻. 何以其数学推导如此简单？其实，测不准原理之深刻在于其物理含意：微观世界中物理量不是用数来表示，而是用自伴算子表示. 至于数学，其作用在于提供了对种种算子进行研究的基础. 例如早就知道矩阵（作为一个算子）的乘法是不可交换的，又例如早就有了普兰舍利定理（这个定理是普兰舍利于 1910 年证明的，早于量子力学的出现），告诉我们只要采取适当的表象（如动量表象），则物理量用自伴算子表示成了很自然的事，所以例如测不准原理等深刻的物理结论，在数学上是水到渠成的事.

然而既然数学上例如普兰舍利定理等等的发现原是与量子力学独立的，则测不准原理也不应该限于只在量子力学中才有. 实际上确实如此. 例如信号处理中对一个信号既可在时域中把它看成 $f(t)$，又可在频域中看成 $\hat{f}(\omega)$. 时间与频率也是典则对偶的，因此也一定有测不准原理：对一个信号，要完全准确地同时确定其旅行时间与其频率是不可能的. 这就是经典的（即非量子的）测不准原理.

## §3 急减函数与缓增广义函数

**1. $S(\mathbf{R}^n)$ 空间** 上两节中讲到 $L^1$ 与 $L^2$ 上的傅里叶级数与傅里叶变换时，遇到的困难都在于函数的可积性上. 由此，例如在傅里叶变换理论中，变换式

$$\hat{f}(\xi)=(Ff)(\xi)=\int_{\mathbf{R}^n} f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x \tag{1}$$

的逆变换式

$$f(x)=(F^{-1}F)(x)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\hat{f}(\xi)\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}x \tag{2}$$

一直无法证明. 所以 (2) 中的 $F^{-1}$ 暂时只是一个形式记号. 傅里叶变换的最基本的性质：乘以自变量 $x$ 或 $\xi$ 与对自变量 $\xi$ 或 $x$ 的求导之关系，也显得很不自然. 多年来，数学家和物理学家都设法推广这些理论，而得到成功与公认的仍是走广义函数理论这条路来讨论傅里叶变换. 上一章中我们已看到广义函数就是试验函数空间的连续线性泛函. 它的基本的"窍门"可以说就是把所遇到的困难都"转移"到试验函数空间去. 我们选取了 $D(\Omega)$ 作为试验函数空间的目的就在于把求导运算的困难仿照分部积分法转移到试验函数上去，而且使"积分号外"的项都自动消失. 为此我们取 $C_0^\infty(\Omega)$ 之元为试验函数，而且为了保证求导运算的"连续性"——即由 $f_n\to f$ 得出 $\partial^\alpha f_n\to\partial^\alpha f$，我们在 $C_0^\infty(\Omega)$ 中引进了一种特别的收敛性即定义 $f_n\to f$ in $D(\Omega)$ 为：所有 $f_n$ 之支集均位于 $\Omega$ 之同一紧子集中，而且对每个固定的 $\alpha$，$\partial^\alpha f_n$ 一致收敛于 $\partial^\alpha f$. 现在问，能不能用同样的方法来讨论 $D(\mathbf{R}^n)$ 上的傅里叶变换，再通过对偶方式把这种变换转移到 $D'(\mathbf{R}^n)$ 上去？当然是可以的. 但是当我们问对 $f(x)\in D(\mathbf{R}^n)$ 按 (1) 作出的 $\hat{f}(\xi)$ 究竟在什么空间中？因此例如在作逆变换 (2)（它与变换 (1) 的差别无非是多了一个因子 $\frac{1}{(2\pi)^n}$，并把 i 变成了 $-$i）时，尽管我们下面就

要看到这仍是可能的，但终究$\hat{f}(\xi)$不再在$D(\mathbf{R}^n)$中了，如何再利用对偶性也就不清楚了．因此要用广义函数理论来处理傅里叶变换时，首先遇到的问题是找一个合适的试验函数空间．当然，由此带来的问题是：这个新空间与原有的$D(\mathbf{R}^n)$是什么关系．

怎样选择试验函数要由所需解决的问题性质决定．考虑到乘以自变量与求导两种运算在傅里叶变换下互变，所以试验函数应具有以下性质：它应该无穷次可微，而且不论求导多少次，或乘以$x$的多少次幂，仍然应该是可积的．因此我们给出下面的

**定义 1** $S(\mathbf{R}^n)$空间——以下简记为$S$——由适合以下条件的$C^\infty$函数$\varphi(x)$构成：对任意重指标$\alpha$与$\beta$，必存在常数$C_{\alpha\beta}$使

$$|x^\beta\partial_x^\alpha\varphi(x)|\leqslant C_{\alpha\beta}. \tag{3}$$

$S$中的函数序列$\{\varphi_k(x)\}$，若对任意固定的重指标$\alpha$与$\beta$，$\{x^\beta\partial_x^\alpha\varphi_k(x)\}$均在$\mathbf{R}^n$上对$x$一致趋向0（不要求对$\alpha$与$\beta$有一致性，即对任意$\varepsilon>0$，均有正整数$K(\varepsilon,\alpha,\beta)$存在，使当$k>K(\varepsilon,\alpha,\beta)$时，$x^\beta\partial_x^\alpha\varphi_k(x)<\varepsilon$，不要求$K$对任意$\alpha,\beta$均相同）就说$\varphi_k(x)\to 0$ in $S$．

$S$函数在无穷远处是急减的，因为由(3)可知不论$N$有多大，恒有

$$|\partial^\alpha\varphi(x)|\leqslant\frac{C_{N,\alpha}}{|x|^N}. \tag{4}$$

即$\partial^\alpha\varphi(x)$之衰减速度高于任意多项式幂之倒数．所以，以后称$S$函数为"急减函数"．一个最常见的例子当然是$e^{-|x|^2}$．

另一方面，任一$C_0^\infty(\mathbf{R}^n)$函数均为急减的：

$$C_0^\infty(\mathbf{R}^n)\subset S.$$

但是这个包含关系只是就集合论意义而言的．若序列$\{\varphi_k(x)\}\to 0$ in $D$，则因 supp $\varphi_k$都含于一个固定的紧集$K$内，所以其中的$x$必有界．此外，$\partial^\alpha\varphi_k(x)$在$K$中一定收敛于0，因此

$$\max|x^\beta\partial^\alpha\varphi_k(x)|\to 0 \text{ in } K. \tag{5}$$

再考虑到在$K$外$x^\beta\partial^\alpha\varphi_k(x)\equiv 0$，所以上面的收敛性其实在$\mathbf{R}^n$上成立，即$\{x^\beta\partial^\alpha\varphi_k(x)\}$在$\mathbf{R}^n$上一致收敛于0：$\{\varphi_k(x)\}\to 0$ in $S$．凡有这样的情况，即有两个各具自己的收敛性的空间$A$和$B$，不但作为集合有$A\subset B$，而且$\{\varphi_k\}\subset A$在$A$的意义下趋于0必导致$\{\varphi_k\}$在$B$的意义下趋于0，就说$A$连续地嵌入$B$中，并记为$A\hookrightarrow B$．这样

$$D(\mathbf{R}^n)\hookrightarrow S.$$

同样也可以看到$S\hookrightarrow L^p(\mathbf{R}^n)$，$1\leqslant p\leqslant\infty$．

现在考虑$S$函数的傅里叶变换(1)．由于$S$函数极好的光滑性与极快的衰减速度，对积分(1)的处理就极为方便：在积分号下求导，使用分部积分法，积分号外各项自动消失等等均不成问题．因此，我们不必证明即可列出$S$中傅里叶变换的性质如下：（逆变换式(2)也暂时承认．下面的函数均指$S$函数．）

(i) 线性性质

$$F(C_1\varphi_1+C_2\varphi_2)=C_1F(\varphi_1)+C_2F(\varphi_2),\ C_1,\quad C_2 \text{ 为复常数}. \tag{6a}$$

$$F^{-1}(C_1\varphi_1+C_2\varphi_2)=C_1F^{-1}(\varphi_1)+C_2F^{-1}(\varphi_2). \tag{6b}$$

(ii) 平移与放大

$$F(\tau_h\varphi)(\xi)=\int_{\mathbf{R}^n}\varphi(x-h)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x=\mathrm{e}^{-\mathrm{i}\xi h}F(\varphi)(\xi).\tag{7a}$$

$$F^{-1}(\tau_h\varphi)(x)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\varphi(\xi-h)\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi=\mathrm{e}^{\mathrm{i}xh}F^{-1}(\varphi)(x).\tag{7b}$$

$$F(\varphi(\lambda\bullet))(\xi)=\int_{\mathbf{R}^n}\varphi(\lambda x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x=\frac{1}{\lambda^n}F(\varphi)\left(\frac{\xi}{\lambda}\right).\tag{8a}$$

$$F^{-1}(\varphi(\lambda\bullet))(x)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\varphi(\lambda\xi)\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi=\frac{1}{\lambda^n}F^{-1}(\varphi)\left(\frac{x}{\lambda}\right).\tag{8b}$$

(iii) **求导与乘以自变量**　凡使用广义函数为工具的文献中，常考虑$\frac{1}{\mathrm{i}}\partial_x$以代替$\partial_x$，并记为$\mathrm{D}_x$. 即$\mathrm{D}_x=\frac{1}{\mathrm{i}}\partial_x$. 这样比较方便. 于是量子力学中的动量算子现在是$\hbar\mathrm{D}_x$.

$$F(\mathrm{D}^\alpha\varphi)(\xi)=\int_{\mathbf{R}^n}(\mathrm{D}^\alpha\varphi)(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x=\xi^\alpha F(\varphi)(\xi).\tag{9a}$$

$$F(x^\beta\varphi)(\xi)=\int_{\mathbf{R}^n}x^\beta\varphi(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x=(-1)^{|\beta|}\mathrm{D}_\xi^\beta F(\varphi)(\xi).\tag{9b}$$

$$F^{-1}(\mathrm{D}_\xi^\alpha\varphi)(x)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}(\mathrm{D}_\xi^\alpha\varphi)(\xi)\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi=(-1)^{|\alpha|}x^\alpha F^{-1}(\varphi)(x),\tag{10a}$$

$$F^{-1}(\xi^\beta\varphi)(x)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\xi^\beta\varphi(\xi)\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi=\mathrm{D}_x^\beta F^{-1}(\varphi)(x).\tag{10b}$$

由这四个式子可见，$S$函数的傅里叶变换与逆傅里叶变换仍是$S$函数，即是说，$F,F^{-1}$：$S\to S$.

不但如此，我们还很容易证明：若$\{\varphi_k\}\to 0$ in $S$，则$F(\varphi_k)(\xi)\to 0$ in $S$，事实上

$$[\xi^\beta\mathrm{D}_x^\alpha F(\varphi_k)](\xi)=(-1)^{|\beta|}[F(\mathrm{D}_x^\beta x^\alpha\varphi_k(\xi))].\tag{11}$$

$\mathrm{D}_x^\beta(x^\alpha\varphi)(x)$自然可以写成$x^\gamma\mathrm{D}_x^\delta\varphi(x)$之线性组合，这时$\gamma\leqslant\alpha,\delta\leqslant\beta$所以这个线性组合是有限的. 现在考虑其中一项. 我们有

$$\begin{aligned}F(x^\gamma\mathrm{D}_x^\delta\varphi_k)(\xi)&=\int_{\mathbf{R}^n}x^\gamma\mathrm{D}_x^\delta\varphi_k(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x\\&=\int_{\mathbf{R}^n}[(1+|x|^2)^Nx^\gamma\mathrm{D}_x^\delta\varphi_k(x)]\frac{\mathrm{e}^{-\mathrm{i}\xi x}}{(1+|x|^2)^N}\mathrm{d}x.\end{aligned}$$

取$N$充分大以至$2N>n$，则$\frac{\mathrm{e}^{-\mathrm{i}\xi x}}{(1+|x|^2)^N}\in L^1(\mathbf{R}^n)$，由于已设$\{\varphi_k\}\to 0$ in $S$，故对任意$\varepsilon>0$，必可找到$K(\varepsilon,\gamma,\delta)$使当$k>K$时，$|(1+|x|^2)^Nx^\gamma\mathrm{D}_x^\delta\varphi_k(x)|<\varepsilon$. 因此

$$\begin{aligned}|F(x^\gamma\mathrm{D}^\delta\varphi_k)(\xi)|&\leqslant\int|(1+|x|^2)^Nx^\gamma D_x^\delta\varphi_k(x)|\frac{|\mathrm{e}^{-\mathrm{i}\xi x}|}{(1+|x|^2)^N}\mathrm{d}x\\&\leqslant\varepsilon\int_{\mathbf{R}^n}\frac{\mathrm{d}x}{(1+|x|^2)^N}=I\varepsilon.\end{aligned}$$

所以$F(x^\gamma\mathrm{D}^\delta\varphi_k)(\xi)$在$\mathbf{R}_\xi^n$上一致趋于0. 代入(11)即知$F(\varphi_k)\to 0$ in $S$. 所以$F:S\to S$也是连

续映射. 同理 $F^{-1}:S\to S$ 也是连续映射. 但是我们还不能说 $F$ 是线性同构,因为(1)与(2)是否互为逆映射尚未证明,$F^{-1}$暂时还只是形式记号. 下面我们就来证明(2)确实是(1)的逆变换.

(iv) **反演公式** 我们来证明(2)确实是(1)的逆变换. 为此我们先来求高斯函数 $e^{-|x|^2/2}=\prod_{j=1}^{n}e^{-x_j^2/2}$ 的傅里叶变换:

$$\int_{\mathbf{R}^n}e^{-|x|^2/2}e^{i\xi x}dx=\prod_{j=1}^{n}\int_{\mathbf{R}}e^{-x_j^2/2}e^{-i\xi_j x_j}dx_j.$$

所以只要计算右方的一个因子

$$\int_{-\infty}^{\infty}e^{-x^2/2}e^{-ix\xi}dx=E(\xi)$$

即可. 上面我们已经说了,$e^{-x^2/2}\in S(\mathbf{R}_x)$,所以其傅里叶变换 $E(\xi)\in S(\mathbf{R}_\xi^n)$. 但是

$$De^{-x^2/2}=ixe^{-x^2/2},$$

所以,利用(9a)与(9b)有

$$\xi E(\xi)=-iD_\xi E(\xi)\quad 或\quad -\xi E(\xi)=\frac{dE(\xi)}{d\xi}.$$

此外,

$$E(0)=\int_{-\infty}^{\infty}e^{-x^2/2}dx=\sqrt{2\pi}.$$

由这两个式子即知

$$E(\xi)=\sqrt{2\pi}e^{-\xi^2/2}.$$

这样我们看到高斯函数 $e^{-x^2/2}$ 的傅里叶变换仍是一个高斯函数,不过因子有一些变化. 对 $E(\xi)$ 再作变换(2),则 $(F^{-1}E)(x)$ 应适合

$$\frac{d}{dx}(F^{-1}E)(x)=-x(F^{-1}E),$$

$$(F^{-1}E)(0)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\sqrt{2\pi}e^{-\xi^2/2}d\xi=1,$$

因此

$$(F^{-1}E)(x)=e^{-x^2/2}.$$

我们只计算了 $e^{-|x|^2/2}$ 的一个因子,把它的 $n$ 个因子全部算完,即得

**引理1** 高斯函数 $e^{-|x|^2/2}$ 的傅里叶变换是 $(2\pi)^{\frac{n}{2}}e^{-|\xi|^2/2}$,从而 $F^{-1}((2\pi)^{\frac{n}{2}}e^{-|\xi|^2/2})=e^{-|x|^2/2}$.

我们时常需要考虑 $f(x)=e^{-a|x|^2/2}\ (a>0)$ 的傅里叶变换,我们不妨把 $f(x)$ 写成 $e^{-|\lambda x|^2/2}$,$\lambda=\sqrt{a}$ 再利用(8a)即有

$$F(e^{-a|x|^2/2})(\xi)=\left(\frac{2\pi}{a}\right)^{\frac{n}{2}}e^{-|\xi|^2/2a}. \tag{12}$$

**定理2** 当 $f(x)\in S$ 时,(2)式定义的变换确是(1)之逆变换,且可写为

$$\hat{\hat{f}}(x) = (2\pi)^n \check{f}(x) = (2\pi)^n f(-x). \tag{13}$$

**证** 回忆一下上节中“证明”$L^2$ 函数之傅里叶变换确是酉变换时,实际上已用到了这里的结论. 当时我们就指出,其中变换积分次序是不合法的. 现在对 $f(x)\in S$,再重复一下这个形式“证明”,我们有

$$(F^{-1}\circ F)f(x) = \frac{1}{(2\pi)^n}\int_{\mathbf{R}_\xi^n} e^{i\xi x}d\xi\int_{\mathbf{R}_y^n} f(y)e^{-i\xi y}dy$$

$$= \frac{1}{(2\pi)^n}\int_{\mathbf{R}_y^n} f(y)dy\int_{\mathbf{R}_\xi^n} e^{i\xi(x-y)}d\xi.$$

如果这个形式证明是合法的,而且又能证到

$$\frac{1}{(2\pi)^n}\int_{\mathbf{R}_\xi^n} e^{i\xi(x-y)}d\xi = \delta(x-y), \tag{14}$$

代入上式就有

$$(F^{-1}\circ F)f(x) = \frac{1}{(2\pi)^n}\int_{\mathbf{R}_\xi^n} f(y)\delta(x-y)dy = f(x),$$

而定理得证. 这个“证明”之所以仍只是形式证明,在于 $e^{i\xi(x-y)}$ 对于 $\xi$ 是不可积的,更不是绝对可积的,所以不能应用富比尼定理以改变积分次序. 对于(14)式,注意到 $F(\delta)(\xi)=1$,而我们想要证明的其实是

$$(F^{-1}\circ F)\delta(x-y) = \delta(x-y), \tag{15}$$

亦即反演公式对于一个很特殊的广义函数是成立的. 这两点其实是一回事:$e^{i\xi(x-y)}$ 是一个急剧振荡的函数(当 $|\xi|\to\infty$ 时),傅里叶级数理论与傅里叶变换理论都在相当大的程度上依赖于这个振荡性质. §1 中讲的黎曼–勒贝格引理,以及它在证明许多基本结果(例如傅里叶级数的狄利克雷定理)中所起的作用,都说明了这种急剧振荡性质起了关键的作用.

如何解决上面提出的矛盾呢? 正如在 §1 中当傅里叶级数收敛性难以证明时,我们引入了种种求和法,例如引入了 $C-1$ 求和以及费耶核. 但是,引入费耶核相当于引进了一个收敛因子以改进收敛性. 如果函数 $f(x)$ 有相当的光滑性(例如适合狄利克雷条件),则这种收敛因子将会加快收敛速度, 如果连这样少的光滑性也没有,则这种收敛因子会导致种种可求和性(如相应于费耶核的 $C-1$ 求和性)以推广收敛性. 所以我们现在也引入一个收敛因子 $\psi(\xi)$ 并考虑

$$\frac{1}{(2\pi)^n}\int_{\mathbf{R}_y^n} dy\int_{\mathbf{R}_\xi^n} f(y)\psi(\varepsilon\xi)e^{i\xi(x-y)}d\xi,$$

这里 $\varepsilon>0$. 我们要来看一下,当 $\varepsilon\to 0$ 时能否得到极限式 $(F^{-1}\circ F)(f)(x)$. 为此,以下两条件是合理的: (i) $\psi(\xi)\in S(\mathbf{R}_\xi^n)$, (ii) $\psi(0)=1$. 然后我们有

$$\frac{1}{(2\pi)^n}\int_{\mathbf{R}_\xi^n} e^{i\xi x}d\xi\int_{\mathbf{R}_y^n} f(y)e^{-i\xi y}dy = \lim_{\varepsilon\to 0}\frac{1}{(2\pi)^n}\int_{\mathbf{R}_y^n}\psi(\varepsilon\xi)e^{i\xi x}d\xi\int_{\mathbf{R}_y^n} f(y)e^{-i\xi y}dy. \tag{16}$$

这时甚至会以为可在积分号下取极限,因为为内层积分是 $\xi$ 的 $S$ 函数而且与 $\varepsilon$ 无关,$\psi(\varepsilon\xi)\to\psi(0)=1$,对每点 $\xi$ 又都成立,然后可以应用勒贝格控制收敛定理. 当然,这样求极限是不行的,

因为 1 并非 $\mathbf{R}_\xi^n$ 上的勒贝格可积函数，不能应用勒贝格控制收敛定理．但是如果注意到对上式右方现在可以自由地应用富比尼定理，即知

$$\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_\xi^n}\psi(\varepsilon\xi)\mathrm{d}\xi\int\limits_{\mathbf{R}_y^n}f(y)\mathrm{e}^{\mathrm{i}\xi(x-y)}\mathrm{d}y=\frac{1}{(2\pi)^n}\iint\limits_{\mathbf{R}_\xi^n\times\mathbf{R}_y^n}\psi(\varepsilon\xi)f(y)\mathrm{e}^{\mathrm{i}\xi(x-y)}\mathrm{d}y\mathrm{d}\xi.$$

令 $\varepsilon\xi=\zeta$，$y-x=\varepsilon z$，上式右方化为

$$\begin{aligned}&\frac{1}{(2\pi)^n}\iint\limits_{\mathbf{R}_z^n\times\mathbf{R}_\zeta^n}\psi(\zeta)f(x+\varepsilon z)\mathrm{e}^{-\mathrm{i}\zeta z}\mathrm{d}\zeta\mathrm{d}z\\&=\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_z^n}f(x+\varepsilon z)\mathrm{d}z\int\limits_{\mathbf{R}_\zeta^n}\psi(\zeta)\mathrm{e}^{-\mathrm{i}\zeta z}\mathrm{d}\zeta\\&=\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_z^n}f(x+\varepsilon z)\hat{\psi}(z)\mathrm{d}z.\end{aligned}$$

现在求极限的过程全转到 $f$ 中去了，而且 $f\in S$，完全可以转到在积分号下求极限，于是代入(16)有

$$\begin{aligned}\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_\xi^n}\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi\int\limits_{\mathbf{R}_y^n}f(y)\mathrm{e}^{\mathrm{i}\xi y}\mathrm{d}y&=\lim_{\varepsilon\to0}\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_z^n}f(x+\varepsilon z)\hat{\psi}(z)\mathrm{d}z\\&=f(x)\cdot\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_z^n}\hat{\psi}(z)\mathrm{d}z.\end{aligned}\tag{17}$$

其实最后一个积分即 $\hat{\psi}(z)$ 之逆傅里叶变换 $\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_z^n}\hat{\psi}(z)\mathrm{e}^{\mathrm{i}\eta z}\mathrm{d}z$ 在 $\eta=0$ 处的值．这里我们要用到 $\hat{\psi}(z)\in S(\mathbf{R}_z^n)$，而这由 $\psi(\xi)\in S\subset(\mathbf{R}_\xi^n)$ 是自然成立的．所以，只要反演公式对 $\psi(\xi)$ 成立，则上式化为 $\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}_z^n}\hat{\psi}(z)\mathrm{d}z=\psi(0)=1$，而定理得证．

但是由引理 1，当 $\psi(z)=\mathrm{e}^{-|\xi|^2/2}$ 时，反演公式确实是成立，故定理得证．

**注** 上面我们是利用了已知高斯函数的傅里叶变换来证明反演公式的，而一旦反演公式已经得知，则对任意 $\psi(\xi)\in S$ 且适合 $\psi(0)=1$ 者，都有 $\frac{1}{(2\pi)^n}\int\limits_{\mathbf{R}^n}\hat{\psi}(z)\mathrm{d}z=\psi(0)=1$，而可以代入(16)式再得出反演公式．这当然不能算作定理的"新证"，因为在这个过程中已经使用了定理 2 了，但是它告诉我们，用任意 $\psi(\xi)$

也是对的．即是说 $K_\varepsilon(x-y)$ 是 $D(\mathbf{R}^n)$ 中的 $\delta$ 序列．但是我们将引入 $S(\mathbf{R}^n)$ 的对偶空间 $S'(\mathbf{R}^n)$，而可以看到 $K_\varepsilon(x-y)$ 也是 $S'(\mathbf{R}^n)$ 中的 $\delta$ 序列．所以，从广义函数角度来看，证明定理 2 的过程就是一个构造 $\delta$ 序列的过程．这和傅里叶级数理论的情况是一样的：为了证明傅里叶级数的收敛性以及可求和性，我们实际上证明的是狄利克雷核、费耶核等，$\left\{\dfrac{\sin\left(N+\frac{1}{2}\right)t}{2\pi\sin\frac{t}{2}}\right\}$，$\left\{\dfrac{\sin^2\left(N+\frac{1}{2}\right)t}{(N+1)\sin^2\frac{t}{2}}\right\}$，$\left\{\dfrac{\sin\left(N+\frac{1}{2}\right)t}{\pi t}\right\}$，…都是 $\delta$ 序列．还有许多其它的 $\delta$ 序列，相应于其它求和法，例如泊松求和等等，而在各自的问题中都很有用．

上面说了 $K_\varepsilon(x-y)\to\delta(x-y)$，所以我们时常就写

$$\delta(x-y)=\frac{1}{(2\pi)^n}\int \mathrm{e}^{\mathrm{i}\xi(x-y)}\mathrm{d}\xi.$$

于是本来没有意义的形式积分现在理解为广义函数的极限，下面还可将把它理解为 1 的逆傅里叶变换，而且把 1 看作一个广义函数，这样理解的积分称为“振荡积分”(在许多文献中振荡积分是用其它方法来定义的，但与这里添加收敛因子的定义是等价的)．这个公式还与其它一些数学理论有关，特别是还可以从量子力学的测不准原理来看待它：粗略地说，$\delta(x)$ 在 $x$ 表象中表示一个粒子的位置完全确定，因此，在 $\xi$ 表象中，它应该完全不确定，而可以用各个方向的平面波 $\mathrm{e}^{\mathrm{i}\xi x}$ 对 $\xi$ 按完全相同的“权”叠加而成，所以我们才有了(14)式(其中令 $y=0$)．这样也就可以理解，为什么 $\delta$ 序列有的并不是如钟型的曲线(正态分布)而有振荡性．现在我们只特别要提醒，在处理含有高频振荡的因子 $\mathrm{e}^{\mathrm{i}\omega\Phi(x,y)}$ $(|\omega|\gg1)$ 的积分时，时常这个积分并不绝对收敛，我们就不要轻易地把绝对值符号放在积分号下去，那样就会使 $|\mathrm{e}^{\mathrm{i}\omega\Phi}|=1$ 而丧失了非常有用的关于相位的信息，而宁可“形式”地把它当作绝对收敛积分，“形式”地应用富比尼定理，再用广义函数或其它方法加以解释．我们正是这样做的．

(v) **酉性质**　还有几个性质都与傅里叶变换作为酉算子有关，因而是极重要的，我们称之为酉性质．

**定理 3**　若 $f,g\in S(\mathbf{R}^n)$，则

(i) $\displaystyle\int\hat{f}(\xi)g(\xi)\mathrm{d}\xi=\int f(x)\hat{g}(x)\mathrm{d}x$，　　(18)

(ii) $\displaystyle\int g(x)\bar{f}(x)\mathrm{d}x=(2\pi)^{-n}\int\overline{\hat{f}(\xi)}\hat{g}(\xi)\mathrm{d}\xi$，　　(19)

(iii) $\widehat{f*g}(\xi)=\hat{f}(\xi)\cdot\hat{g}(\xi)$，$\dfrac{1}{(2\pi)^n}(\hat{f}*\hat{g})(\xi)=\quad f\cdot g$

性的，后者则是共轭线性的. 欧几里得配对是用来定义转置算子 $F^*$ 的.

**注 2** (19)可以写为

$$(f,g)=(2\pi)^{-n}(\hat{f},\hat{g}),$$

所以傅里叶变换又是等距的. 但通常讲等距算子是对希尔伯特空间而言，现在 $S(\mathbf{R}^n)$并非希尔伯特空间. 所以说 $F$ 是等距的，或者是酉算子，都只是方便的说法，而在 $L^2(\mathbf{R}^n)$中它才有精确的意义. 借用这种方便的说法，(19)也称为帕塞瓦尔关系式，其实也可称为普兰舍利关系式，不过一般说来，只当 $f=g$ 时，才称为**普兰舍利定理**.

**注 3** (20)是卷积公式. 当 $f,g\in S$ 时

$$(f*g)(x)=\int_{\mathbf{R}^n} f(t)g(x-t)\mathrm{d}t.$$

显然积分是收敛的，而且因为可以在积分号下求导，所以$(f*g)(x)\in C^\infty(\mathbf{R}^n)$也是明显的，而且

$$\partial_x^\alpha(f*g)(x)=\int_{\mathbf{R}^n} f(t)\partial_x^\alpha g(x-t)\mathrm{d}t.$$

余下只要考虑 $x^\beta\partial_x^\alpha(f*g)(x)$，因为我们只在$|x|\gg 1$ 处考虑它，不妨用$(1+|x|)^{|\beta|}$来代替$|x|$，于是

$$|x^\beta\partial_x^\alpha(f*g)(x)|\leqslant\int_{\mathbf{R}^n}(1+|x|)^{|\beta|}|f(t)\partial_x^\alpha g(x-t)|\mathrm{d}t.$$

因为$|x|\leqslant|t|+|x-t|$，而$|\beta|\geqslant 0$，上式右方可以用多项式定理化为$|t|^\gamma|f(t)|\cdot|x-t|^\delta\ |\partial_x^\alpha g(x-t)|$之有限线性组合. 这里$|\gamma|+|\delta|\leqslant|\beta|$. 由于 $f,g\in S$，即知上式右方是一个绝对收敛积分，其值$\leqslant C_{\alpha,\beta}$. 这样即知$(f*g)(x)\in S(\mathbf{R}^n)$，而(20)就容易证明了.

**2. $S'(\mathbf{R}^n)$广义函数** 这就是急减函数空间 $S(\mathbf{R}^n)$的对偶.

**定义 2** $S(\mathbf{R}^n)$上的线性连续泛函称为缓增广义函数，它们构成一个复系数线性空间，记为$S'(\mathbf{R}^n)$.

下面我们举一些例子.

$L^p(\mathbf{R}^n)\hookrightarrow S'(\mathbf{R}^n)\ (1\leqslant p\leqslant+\infty)$，因为若 $f(x)\in L^p(\mathbf{R}^n)$，对 $\varphi(x)\in S(\mathbf{R}^n)$，$\int_{\mathbf{R}^n} f(x)\varphi(x)\mathrm{d}x$显然有定义，而且由赫尔德不等式

$$\left|\int_{\mathbf{R}^n} f(x)\varphi(x)\mathrm{d}x\right|\leqslant\left[\int_{\mathbf{R}^n}|f(x)|^p\mathrm{d}x\right]^{\frac{1}{p}}\left[\int_{\mathbf{R}^n}|\varphi(x)|^q\mathrm{d}x\right]^{\frac{1}{q}},$$

$$\frac{1}{p}+\frac{1}{q}=1,$$

由此可知若有一串 $\varphi_k(x)\to 0$ in $S(\mathbf{R}^n)$，必有 $\int_{\mathbf{R}^n} f(x)\varphi_k(x)\mathrm{d}x\to 0$，所以它是一个连续线性泛函，这样我们得到 $L^p(\mathbf{R}^n)\subset S'(\mathbf{R}^n)$. 为了证明它是连续嵌入，设有一串 $f_k(x)\in L^p(\mathbf{R}^n)$，$f_k(x)\to f(x)$ in $L^p(\mathbf{R}^n)$. 仍旧用上面的赫尔德不等式，即知 $f_k(x)\to f(x)$ in $S'(\mathbf{R}^n)$，($S'(\mathbf{R}^n)$中的收敛性定义，下面马上就要介绍)，所以这里的嵌入是连续嵌入.

但是 $S'(\mathbf{R}^n)$中的元不一定都是可积函数. 实际上,若 $f(x)$是可积函数,甚至只是局部可积函数,但在∞附近适合以下的缓增条件:存在正整数 $N$ 以及常数 $C_N$,使得

$$|f(x)|\leqslant C_N(1+|x|)^N.$$

(或者写得更简单一些,写作 $f(x)=O(1+|x|)^N$,即是说 $f(x)$之增长速度不超过 $N$ 次多项式的增长速度.)由于 $\varphi(x)\in S(\mathbf{R}^n)$是急减的,因此例如有

$$|\varphi(x)|\leqslant\frac{A}{(1+|x|)^{N+n+1}}.$$

所以

$$\left|\int_{\mathbf{R}^n}f(x)\varphi(x)\mathrm{d}x\right|\leqslant\int_{\mathbf{R}^n}\frac{C_NA}{(1+|x|)^{n+1}}\mathrm{d}x$$

仍是有意义的.

但是我们不要以为 $S'(\mathbf{R}^n)$中的一切函数都必定是缓增的. 例如我们知道 $\mathrm{e}^{|x|}$ 是指数增长的,它不是缓增的,在 $n=1$ 的情况,$[\cos(\mathrm{e}^x)]'=-\mathrm{e}^x\sin(\mathrm{e}^x)$,也不是缓增的,但是它是 $S'(\mathbf{R}^n)$中的元. 因为对于 $\varphi(x)\in S(\mathbf{R}^1)$.

$$\langle[\cos(\mathrm{e}^x)]',\varphi(x)\rangle=-\int_{-\infty}^{\infty}[\cos(\mathrm{e}^x)]\varphi'(x)\mathrm{d}x,$$

它显然是 $S(\mathbf{R}^1)$上的连续线性泛函.

另一方面,由 $D(\mathbf{R}^n)\subsetneq S(\mathbf{R}^n)$,所以每一个 $S'(\mathbf{R}^n)$广义函数一定是 $D'(\mathbf{R}^n)$广义函数. 因此

$$S'(\mathbf{R}^n)\subset D'(\mathbf{R}^n).$$

我们要证明实际上有 $S'(\mathbf{R}^n)\subsetneq D'(\mathbf{R}^n)$. 为此,我们先要讲一下什么是 $S'(\mathbf{R}^n)$中的收敛性. 设有 $f_0(x)\in S'(\mathbf{R}^n)$,而且有一串 $f_k(x)\in S'(\mathbf{R}^n)$,$f_k(x)\to f_0(x)$ in $S'(\mathbf{R}^n)$是什么意义? 为此,我们只需了解 $f_k(x)-f_0(x)\to 0$ in $S'(\mathbf{R}^n)$是什么意义即可. 为此我们仿照 $D'(\mathbf{R}^n)$中趋于 0 的定义给出:

**定义 3**　$S'(\mathbf{R}^n)$广义函数 $f_k(x)\to 0$ in $S'(\mathbf{R}^n)$即指对任意 $\varphi(x)\in S(\mathbf{R}^n)$,均有 $\langle f_k(x),\varphi($

函数 $\chi(x)$，使在 $K$ 附近 $\chi(x)=1$，于是对任一 $\varphi\in S$，$(\chi\varphi)(x)\in D(\mathbf{R}^n)$，而我们对 $\varphi\in S(\mathbf{R}^n)$ 定义

$$\langle f,\varphi\rangle=\langle f,\chi\varphi\rangle.$$

很容易验证，这个定义与 $\chi$ 之选择无关，这样 $D'(\mathbf{R}^n)$中的 $f$ 现在拓展成$S'(\mathbf{R}^n)$中之元.

特别重要的是，可以证明，$C_0^\infty(\mathbf{R}^n)$在 $D'(\mathbf{R}^n)$中稠密，所以在 $S'(\mathbf{R}^n)$中稠密. 即是说，对任意 $f\in S'(\mathbf{R}^n)$，必可找到一个 $C_0^\infty(\mathbf{R}^n)$序列 $f_k(x)$，使

$$f_k(x)\to f(x)\quad \text{in}\quad S'(\mathbf{R}^n).$$

这个定理很重要，还因为它把 $D'$广义函数的性质，大都移到 $S'$广义函数中来了. 我们不来一一列举，只举出其中最重要的几个. 以下我们均设 $f\in S'(\mathbf{R}^n)$，$\varphi(x)\in S(\mathbf{R}^n)$

(i) **平移**　$\tau_h f=f(x-h)$的定义是：

$$\langle\tau_h f,\varphi\rangle=\langle f,\tau_h\varphi\rangle. \tag{21}$$

(ii) **反射**　$\check{f}(x)=f(-x)$的定义是：

$$\langle\check{f},\varphi\rangle=\langle f,\check{\varphi}\rangle. \tag{22}$$

(iii) **求导**之定义现在成为

$$\langle \mathrm{D}^\alpha f,\varphi\rangle=(-1)^{|\alpha|}\langle f,\mathrm{D}^\alpha\varphi\rangle. \tag{23}$$

注意，由于我们引入 $S$ 与 $S'$是为了傅里叶变换理论的需要，所以必须考虑复值函数，故定义 $S$ 上的线性泛函时需要应用埃尔米特配对$(f,\varphi)$，它对后一个“变元”$\varphi$ 是共轭线性的：

$$(\lambda f,\varphi)=\lambda(f,\varphi),\quad (f,\lambda\varphi)=\bar{\lambda}(f,\varphi).$$

而在讨论 $D(\Omega)$与 $D'(\Omega)$时，如果限于讨论实值函数，则应该用欧几里得配对$\langle f,\varphi\rangle$，这时有

$$\langle\lambda f,\varphi\rangle=\langle f,\lambda\varphi\rangle=\lambda\langle f,\varphi\rangle\quad(\lambda\ 为实数).$$

仿此，在讨论求导运算时，我们一般都引用 $\mathrm{D}=\dfrac{1}{\mathrm{i}}\partial$，从而有定义(23). 而在 $D(\Omega)$和 $D'(\Omega)$中则有

$$\langle\partial^\alpha f,\varphi\rangle=(-1)^{|\alpha|}\langle f,\partial^\alpha\varphi\rangle.$$

但是“乘以函数”的运算在 $D'(\Omega)$和 $S'(\mathbf{R}^n)$中是不一样的. 对于 $D'(\Omega)$的情况，任给

今设 $\varphi\in S(\mathbf{R}^n)$，于是 $\hat{\varphi}$ 也是 $S$ 函数，若 $f\in S'(\mathbf{R}^n)$，则因傅里叶变换 $F:S(\mathbf{R}^n)\to S(\mathbf{R}^n)$ 是线性同构，对任一个给定的 $\hat{\varphi}$ 只有一个 $\varphi$ 与之相应，因此 $\langle f,\hat{\varphi}\rangle$ 是 $\varphi$ 的线性泛函. 又因 $F$ 与 $F^{-1}:S(\mathbf{R}^n)\to S(\mathbf{R}^n)$，都是连续的，故若有一串 $\varphi_k\in S(\mathbf{R}^n)$，但 $\varphi_k\to 0$ in $S(\mathbf{R}^n)$，则 $\hat{\varphi}_k\to 0$ in $S(\mathbf{R}^n)$，因此 $\langle f,\hat{\varphi}_k\rangle\to 0$，即是说 $\langle f,\hat{\varphi}_k\rangle$ 其实是 $\varphi\in S(\mathbf{R}^n)$ 上的连续线性泛函，因此必存在唯一的 $S'(\mathbf{R}^n)$ 之元 $g\in S'(\mathbf{R}^n)$，使 $\langle f,\hat{\varphi}\rangle=\langle g,\varphi\rangle$. 因此我们给出

**定义 4** 上述 $g\in S'(\mathbf{R}^n)$ 称为 $f\in S'(\mathbf{R}^n)$ 之傅里叶变换，并记为 $\hat{f}$. 因此 $S'(\mathbf{R}^n)$ 广义函数 $f$ 之傅里叶变换 $\hat{f}$ 仍为一 $S'(\mathbf{R}^n)$ 广义函数而定义如下：

$$\langle f,\hat{\varphi}\rangle=\langle\hat{f},\varphi\rangle. \tag{24}$$

显然，$F:f\to\hat{f}$ 是连续映射.

**注** 我们是以欧几里得配对作为定义 4 的基础的，现在 $f$ 与 $\hat{f}$ 均可能为复值函数，故本来应该采用埃尔米特配对，但因埃尔米特配对 $(f,\varphi)$ 对后一变元是共轭线性的，故对复常数 $C_1$ 与 $C_2$ 应有 $(f,C_1\varphi_1+C_2\varphi_2)=\overline{C}_1(f,\varphi_1)+\overline{C}_2(f,\varphi_2)$. 这时，定义 4 需要作一些改动，我们这里就不讲了.

现在回到 $S'(\mathbf{R}^n)$ 的子空间 $L^p$. 先看 $p=1$ 的情况. 于是设 $f(x)\in L^1(\mathbf{R}^n)$，由前所述

$$\begin{aligned}\langle f,\hat{\varphi}\rangle&=\int_{\mathbf{R}^n}f(x)\hat{\varphi}(x)\mathrm{d}x=\int_{\mathbf{R}^n}f(x)\mathrm{d}x\int_{\mathbf{R}^n}\varphi(\xi)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}\xi\\&=\iint_{\mathbf{R}^n\times\mathbf{R}^n}f(x)\varphi(\xi)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x\mathrm{d}\xi\\&=\int_{\mathbf{R}^n}\varphi(\xi)\mathrm{d}\xi\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x,\ \varphi\in S(\mathbf{R}^n).\end{aligned}$$

因为 $f(x)\in L^1(\mathbf{R}^n)$，$\varphi(\xi)\in S(\mathbf{R}^n)$，可以应用富比尼定理而得到最后一个等式. 这个等式我们这样来看：由 $f\in L^1$ 作为 $S'(\mathbf{R}^n)$ 之元的傅里叶变换(变换的结果我们仍记为 $\tilde{f}$)之定义，即由积分 $\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x$ 生成的 $S'(\mathbf{R}^n)$ 之元. 因此

$\tilde{f}$ 作为 $S(\mathbf{R}^n)$ 上之连续线性泛函

$$=\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x\ \text{所生成的}\ S(\mathbf{R}^n)\ \text{上的连续线性泛函}.$$

注意，我们在上一节就已讲过 $\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x$ 作为一个积分是绝对收敛的，因此 $\int_{\mathbf{R}^n}f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x$ 确实是一个古典意义下的函数，不过我们尽管知道它是连续的，在 $|x|\to+\infty$ 时为 0，却不能断定它仍是 $L^1(\mathbf{R}^n)$ 函数(且有反例. 当然可以肯定它是局部可积函数)，把上面式子中的话讲准确些就是：

$f$(作为 $S'(\mathbf{R}^n)$ 之元)的 $S'(\mathbf{R}^n)$ 傅里叶变换 $\tilde{f}$

$= f$ 之古典的傅里叶变换（我们记为 $\hat{f}$）生成的 $S'(\mathbf{R}^n)$ 之元.

在这个意义下我们说 $S'(\mathbf{R}^n)$ 傅里叶变换是 $L^1(\mathbf{R}^n)$ 傅里叶变换的推广. 并且不必再多用一个记号 $\tilde{f}$ 而都记作 $\hat{f}$.

其次若 $f \in L^2(\mathbf{R}^n)$，这时若令

$$f_N(x) = \begin{cases} f(x), & |x_j| \leqslant N,\ j = 1,2,\cdots,n, \\ 0, & \text{其它 } x, \end{cases}$$

于是 $f_N(x) \to f(x) \quad \text{in} \quad L^2(\mathbf{R}^n)$. 按古典的傅里叶变换理论，古典的傅里叶变换 $\hat{f}_N(\xi)$ 在 $L^2(\mathbf{R}^n_\xi)$ 中有极限，这就是 $\hat{f}(\xi)$，它仍是 $L^2(\mathbf{R}^n_\xi)$ 函数. 普兰舍利定理告诉我们，古典的傅里叶变换是一个酉算子.

现在转到 $S'(\mathbf{R}^n)$. 首先可见

$$\langle \hat{f}_N, \varphi \rangle = \langle f_N, \hat{\varphi} \rangle.$$

于是仍照上面的讲法 $\tilde{f}_N = \hat{f}_N$. 但当 $N \to \infty$ 时，$f_N \to f \quad \text{in} \quad L^2(\mathbf{R}^n)$，所以 $\hat{f}_N \to \hat{f} \quad \text{in} \quad L^2(\mathbf{R}^n)$. 不过若把 $f_N$ 作为 $S'(\mathbf{R}^n)$ 之元，$f_N \to f$ 必导致 $\tilde{f}_N$ 在 $S'(\mathbf{R}^n)$ 中有极限. 由定义容易看出，$\tilde{f}_N$ 的 $S'(\mathbf{R}^n)$ 极限即 $f$ 作为 $S'(\mathbf{R}^n)$ 之元之傅里叶变换，记作 $\tilde{f}$. 于是我们又得到

$f$（作为 $S'(\mathbf{R}^n)$ 之元）的 $S'(\mathbf{R}^n)$ 傅里叶变换 $\tilde{f}$

$=$（$f$ 之古典的傅里叶变换）$\hat{f}$ 所生成的 $S'(\mathbf{R}^n)$ 之元.

所以我们也说 $S'(\mathbf{R}^n)$ 傅里叶变换是 $L^2(\mathbf{R}^n)$ 的傅里叶变换的推广，而且我们仍使用同一记号.

从这样的讨论中，读者可能会有一个感觉：至少在 $L^1(\mathbf{R}^n)$，$L^2(\mathbf{R}^n)$ 情况下，广义函数理论并没有给我们任何新的信息. 确实如此，例如对于 $f \in L^1(\mathbf{R}^n)$ 时的 $\hat{f}(\xi)$，除了我们原来就已知道的连续性以及在 $\infty$ 处为 0 以外，没有任何新结果. 那么，这种推广有什么意义呢？我们暂时把这个问题放在一边，先来看一下 $\hat{f}$ 的性质.

对于 $S(\mathbf{R}^n)$ 之傅里叶变换我们有前面讲到的线性、平移与放大、求导与乘以自变量等性质，即前面讲的 (6) ~ (10) 诸式，对于 $S'(\mathbf{R}^n)$ 也是一样，我们可以把它们概括为

**定理 5** $S'(\mathbf{R}^n)$ 上的傅里叶变换 $F: S'(\mathbf{R}^n) \to S'(\mathbf{R}^n)$ 有以下性质：设 $f \in S'(\mathbf{R}^n)$，则有

(i) 线性性质：傅里叶变换是线性变换；

(ii) 平移与放大：

$$\widehat{(\tau_h f)}(\xi) = \mathrm{e}^{-\mathrm{i}\xi h} \hat{f}(\xi), \tag{25a}$$

$$\widehat{f(\lambda \cdot)}(\xi) = \frac{1}{\lambda^n} \hat{f}\left(\frac{\xi}{\lambda}\right); \tag{25b}$$

(iii) 求导与乘以自变量：

$$\widehat{(\mathrm{D}^\alpha f)}(\xi) = \xi^\alpha \hat{f}(\xi), \tag{26a}$$

$$\widehat{(x^\beta f)}(\xi) = (-1)^{|\beta|} \mathrm{D}_\xi^\beta \hat{f}(\xi). \tag{26b}$$

关于逆变换是否也有如上的性质，上面都没有讲，因为下面讲了反演公式以后就清楚了．

我们再来看一下逆变换的形式定义(2)．当然，对于 $S(\mathbf{R}^n)$之元它确实是一个很“规矩”的积分．从形式上看，它与傅里叶变换除了相差一个因子$\frac{1}{(2\pi)^n}$以外，只不过“相位”$\mathrm{i}\xi x$ 差了一个符号．因此，只要对 $x$ 再作一个反射$x\to -x$，(2)就成了(1)(还要加一个因子$\frac{1}{(2\pi)^n}$)，即是说 $f\to\hat{f}\to\hat{\hat{f}}\to$乘以$(2\pi)^{-n}\to$反射$\to f$．所以，在 $S(\mathbf{R}^n)$中的反演公式可以写成

$$\hat{\hat{f}}=(2\pi)^n\check{f}.$$

现在我们要对 $S'(\mathbf{R}^n)$证明同样的结果：

**定理 6** $F$ 与$F^{-1}$都是 $S'(\mathbf{R}^n)$到 $S'(\mathbf{R}^n)$的连续的线性同构，而且

$$\hat{\hat{f}}=(2\pi)^n\check{f},\quad f\in S'(\mathbf{R}^n).\tag{27}$$

**证** $F$ 是连续的线性映射：$S'(\mathbf{R}^n)\to S'(\mathbf{R}^n)$已经明白．现设 $f\in S'(\mathbf{R}^n)$，对于任意的 $\varphi\in S(\mathbf{R}^n)$有

$$\langle\hat{\hat{f}},\varphi\rangle=\langle f,\hat{\hat{\varphi}}\rangle=\langle f,(2\pi)^n\check{\varphi}\rangle=\langle(2\pi)^n\check{f},\varphi\rangle.$$

于是(27)得证．因此 $F^{-1}$是由三个由 $S'(\mathbf{R}^n)$到 $S'(\mathbf{R}^n)$的连续线性映射合成(两次傅里叶变换，再加一次反射)，所以自己也是连续线性变换．至于傅里叶变换已经由定理 2 得知它是连续线性映射．总之，$F,F^{-1}:S'(\mathbf{R}^n)\to S'(\mathbf{R}^n)$都是连续的线性同构．定理证毕．

**推论 7** 对于逆傅里叶变换有

(i) 线性性质：自明；

(ii) 平移与放大：

$$F^{-1}(\tau_h f)(x)=\mathrm{e}^{\mathrm{i}xh}F^{-1}(f)(x),\tag{28a}$$

$$F^{-1}(f(\lambda\cdot))(x)=\frac{1}{\lambda^n}F^{-1}(f)\left(\frac{x}{\lambda}\right);\tag{28b}$$

(iii) 求导与乘以自变量

$$F^{-1}(\mathrm{D}_\xi^\alpha f)(x)=(-1)^{|\alpha|}x^\alpha F^{-1}(f)(x),\tag{29a}$$

$$F^{-1}(\xi^\beta f)(x)=\mathrm{D}_x^\beta F^{-1}(\hat{f})(x).\tag{29b}$$

这些推论我们都不来证明了．

现在要问关于 $S(\mathbf{R}^n)$中的傅里叶变换的定理 3，对 $S'(\mathbf{R}^n)$的傅里叶变换有无对应定理？其实(18)式现在就是以欧几里得配对为基础的定义 4．也就是如何求 $F$ 的转置算子．(19)式则可以说就是以埃尔米特配对为基础，求 $F$ 的伴算子．(20)比较困难，它涉及如何定义两个广义函数 $f$ 与 $g$ 的卷积．

广义函数的卷积是一个困难的问题，我们这里不打算多讨论它，但想从形式推导着手指出其困难所在．如果 $f(x),g(x)$作为 $D'(\mathbf{R}^n)$之元其卷积仍可用 $(f*g)(x)=\int f(t)g(x-t)\mathrm{d}t$

来定义，则对于 $\varphi(x)\in D(\mathbf{R}^n)$ 应有

$$\begin{aligned}\langle f*g,\varphi\rangle&=\iint f(t)g(x-t)\varphi(x)\mathrm{d}t\mathrm{d}x\\&=\iint f(t)g(y)\varphi(t+y)\mathrm{d}y\mathrm{d}x\\&=\int f(t)\mathrm{d}t\int g(y)\varphi(t+y)\mathrm{d}y\\&=\langle f(t),\langle g(y),\varphi(t+y)\rangle\rangle.\end{aligned}$$

现在的问题是：尽管 $\varphi(t+y)$ 作为 $t+y$ 的函数具有紧支集，例如 $|t+y|\leqslant R$，但对任意变动的 $t$，$\varphi(t+y)$ 作为 $y$ 的函数不一定有紧支集. 因此 $\langle g(y),\varphi(t+y)\rangle$ 作为 $t$ 的函数不一定在 $D(\mathbf{R}_t^n)$ 中，所以上式最后一项不好定义. 同样，若 $f,g\in S'(\mathbf{R}^n)$，$\varphi\in S(\mathbf{R}^n)$，我们也不一定知道 $\langle g(y),\varphi(t+y)\rangle\in S'(\mathbf{R}_t^n)$.

再说，如果能够定义 $(f*g)(x)$ 了，而且得出 $(\widehat{f*g})(\xi)=\hat{f}(\xi)\cdot\hat{g}(\xi)$（本来，定义卷积很重要的目的就是为了证明这个公式），则两个广义函数 $\hat{f}(\xi)$ 与 $\hat{g}(\xi)$ 的乘积也就可以定义了. 我们在第四章中就说了一般而言是不能定义广义函数的乘积的，因此也就会怀疑，一般而言是不能定义两个广义函数的卷积的. 为此应该对 $f$ 与 $g$ 加上一些条件，例如规定二者之一有紧支集. 在本书中，关于紧支集广义函数的我们未作专门论述. 这样一来许多问题只好请读者去查广义函数的专门著作了.

最后关于反演公式还要讲几句话，上面我们已讲了，$L^1(\mathbf{R}^n)\hookrightarrow S'(\mathbf{R}^n)$，而且 $L^1$ 函数的傅里叶变换 $\int_{\mathbf{R}^n} f(x)\mathrm{e}^{-\mathrm{i}\xi x}\mathrm{d}x$，不论按古典意义去讨论还是按 $S'(\mathbf{R}^n)$ 意义去理解都是一致的. 按古典意义去讨论时，我们知道 $f(x)=\frac{1}{(2\pi)^n}\int_{\mathbf{R}^n}\hat{f}(\xi)\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi$ 一般是不成立的，因为 $\hat{f}(\xi)$ 不一定可积，所以上式一般而言是不对的. 那么 $S'(\mathbf{R}^n)$ 理论是不是可以给我们什么帮助呢？很遗憾，帮助不大. $S'(\mathbf{R}^n)$ 理论只能告诉我们 $\hat{f}(\xi)\in S'(\mathbf{R}^n)$，也许我们愿意接受上式作为一个形式积分，正如同我们愿意接受 $\delta(x)=\frac{1}{(2\pi)^n}\int\mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi$ 那样，也许我们还可以把它解释为振荡积分，但是 $L^1$ 函数的傅里叶变换究竟有些什么特性足以刻画它，这个问题是无法回避的. $L^2$ 的情况好一些：当 $f\in L^2$ 时，$\hat{f}$ 也是 $L^2$ 函数，即当用 $f_N$ 在 $L^2$ 中去逼近 $f$ 时，$\hat{f}_N$ 必有一个 $L^2$ 极限即为 $\hat{f}$（不过，$f_N$ 是有紧支集的，$\hat{f}_N$ 则一定没有紧支集，它是一类增长性受到严格限制的整函数）. 至于 $S'(\mathbf{R}^n)$ 中的其它元素问题就更多了. 有一种错觉，以为有了广义函数理论，傅里叶积分的任何困难都迎刃而解了. 事实当然不是如此，有了 $S'(\mathbf{R}^n)$ 理论

宁可说，大约二百年前傅里叶所开创的数学分支——现在通称为调和分析——一直在蓬勃发展．本章的目的只是为读过微积分的读者提供一个简明的介绍．

我们还是想就 $L^1$ 与 $L^2$ 中傅里叶变换的反演公式说几句话．$L^1$ 与 $L^2$ 函数都是 $S'(\mathbf{R}^n)$之元．因此$\hat{\hat{f}} = (2\pi)^n \check{f}$都是成立的．当 $f\in L^1$ 时，$\hat{f}(\xi) = \int f(x)e^{-i\xi x}dx$ 只是一个连续函数，而且 $\xi\to\infty$时趋于 0．因此虽然仍是一个 $S'(\mathbf{R}^n)$广义函数，其傅里叶变换却不能按古典的勒贝格积分来处理，但若$\hat{f}(\xi)\in L^1(\mathbf{R}^n_\xi)$，则

$$\hat{\hat{f}}(x) = \int \hat{f}(\xi)e^{-i\xi x}dx = (2\pi)^n\cdot\frac{1}{(2\pi)^n}\int \hat{f}(\xi)e^{i\xi y}dy\Big|_{y=-x}.$$

所以，$\hat{\hat{f}}(x) = (2\pi)^n\check{f}(x)$就意味着

$$\frac{1}{(2\pi)^n}\int \hat{f}(\xi)e^{i\xi x}dx = f(x).$$

此式不论在 $S'$意义下还是在勒贝格可积函数意义下都对．所以有

**推论 8**　若 $f(x)\in L^1(\mathbf{R}^n)$，且$\hat{f}(\xi)\in L^1(\mathbf{R}^n)$，则以下的反演公式成立：

$$f(x) = \frac{1}{(2\pi)^n}\int \hat{f}(\xi)e^{i\xi x}dx.$$

$L^2$ 函数的情况比较简单，若 $f\in L^2$，则$\hat{f}$无论按$S'$意义还是按$L^2$ 意义均为

$$\hat{f}(\xi) = \int f(x)e^{i\xi x}dx = \underset{N\to\infty}{\text{l.i.m.}}\int f_N(x)e^{i\xi x}dx.$$

但$\hat{f}(\xi)$一定仍是 $L^2$ 函数，所以一定可以再作傅里叶变换，而且不论是在 $S'$意义下还是在$L^2$ 意义下都相同：

$$\hat{\hat{f}}(x) = \underset{N\to\infty}{\text{l.i.m.}}\int (\hat{f})_N(\xi)e^{i\xi x}d\xi = \int \hat{f}(\xi)e^{-i\xi x}d\xi.$$

由于$\hat{\hat{f}}(x) = (2\pi)^n\check{f}(x)$，所以

$$\int \hat{f}(\xi)e^{-i\xi x}d\xi = \hat{\hat{f}}(x) = (2\pi)^n f(-x).$$

于是有

**推论 9**　若 $f(x)\in L^2(\mathbf{R}^n)$．则不论在 $S'$意义下还是在$L^2$ 意义下均有

$$f(x) = \frac{1}{(2\pi)^n}\int \hat{f}(\xi)e^{i\xi x}d\xi.$$

最后余下一个问题，即 $S'(\mathbf{R}^n)$乘子问题．我们要问，什么样的 $C^\infty$ 函数 $a(x)$与任一个 $S'(\mathbf{R}^n)$广义函数 $f(x)$之乘积仍是 $S'(\mathbf{R}^n)$广义函数？这类 $a(x)$就称为 $S'$乘子．我们知道，$a(x)f(x)$应由下式定义：当 $a\in C^\infty(\mathbf{R}^n)$，$f\in S$

需要 $a(x)$ 在 $\infty$ 处增长不能太急，例如 $a(x)=\mathrm{e}^{|x|^2}$ 一定不行．因为 $\varphi(x)=\mathrm{e}^{-|x|^2/2}$ 显然是 $S(\mathbf{R}^n)$ 函数，读者自己可以证明，但 $a\varphi=\mathrm{e}^{|x|^2/2}$ 显然不在 $S(\mathbf{R}^n)$ 中，这都是因为 $a(x)$ 是指数增长的原故，于是我们有

**定理 10** 设 $a(x)\in C^\infty(\mathbf{R}^n)$，对任一 $\varphi(x)\in S(\mathbf{R}^n)$，均有 $a(x)\varphi(x)\in S(\mathbf{R}^n)$ 的充分必要条件是：对任意重指标 $\alpha$，均存在一个多项式 $P_\alpha(x)$，使

$$|\partial^\alpha a(x)|\leqslant|P_\alpha(x)|. \tag{31}$$

以下我们记这种 $a(x)$ 所成的空间为 $O_M(\mathbf{R}^n)$．

于是我们来看 $S'$ 乘子空间：$\{a(x)\in C^\infty$，对一切 $f\in S'(\mathbf{R}^n)$，$a(x)f\in S'(\mathbf{R}^n)\}$，我们有

**定理 11** $S'(\mathbf{R}^n)$ 乘子空间即 $O_M(\mathbf{R}^n)$．

这里的几个定理我们都不给证明了．

# 第六章　再论微积分的基础

## §1　实数理论

微积分的理论基础是在微积分发展到相当程度以后才奠立起来的.我们在第一章一开始就指出,微积分既然是变量的数学,而变量又是一个极困难的概念,要想明确地理解它而不出逻辑上的矛盾,必须要等到具体的数学结果积累到相当丰富,把问题的各个方面都暴露到相当程度才行.所以,尽管在希腊时代就已经认识到诸如连续性、运动、无穷小等等是多么困难的概念,甚至欧多克萨斯关于比例的理论,关于不可通约量的理论与现代的实数理论颇为相近;在欧几里得关于圆面积的穷竭性理论中也看到了如何把极限问题化为不等式问题.但我们究竟不能说这些古希腊的著作就是今天通用的微积分教本所介绍的理论的前身.因为一些最根本的概念一直到19世纪中晚期才弄明白.而在希腊时代,数学还与种种思辨性的议论分不开.不仅毕达哥拉斯是这样,芝诺和亚理士多德更是这样.可是,科学(数学也是其一部分)还是在沿着探索宇宙的规律的大道上阔步前进.到了19世纪,这些问题在当时的水平上弄得相当清楚了.数学,特别是微积分,在说明宇宙的规律取得何等伟大的成就,相信前五章的内容已经足够给读者一个印象.既然如此,又何必回到历史的陈迹中,再去追忆当年呢?问题在于,数学还在前进.今天已不再是达朗贝尔“前进吧,您会得到信心”的时代.其实,自从所谓微积分的基础得到奠立时起,整个数学又陷入更深刻的困难之中.这就是集合论的建立.论时间是19世纪七八十年代,主要的代表人物是康托尔(Georg Cantor),而且他的起点正是黎曼,狄利克雷关于三角级数的工作,我们在第五章一开始就介绍了一点点有关情况.今天我们再回到微积分的基础是因为,弄明白了这个基础,就发现它含有许多“成分”——用布尔巴基的语言来说是许多数学构造或称为 architecture.这些构造甚至规定了下一步数学将怎样走法.因此,本章的目的是从介绍微积分的基础使读者看到这些构造是些什么.其实在前几章中我们已在许多具体问题上与它们打过交道了.所以,我们希望是从再论微积分的基础中帮助读者能与更新、更现代的数学打交道.

**1. 有理数系怎样扩大成实数系**　为微积分奠立基础的工作现在都说是其算术化的过程.其原因在于极限理论过去实际上是以几何直观为基础的.例如,我们说$\lim\limits_{x\to a}f(x)=b$ 时,心目中总会想到当 $x$ 连续地沿 $x$ 轴趋向 $a$ 时,$f(x)$沿另一条轴趋向 $b$ 点.可是如果在 $x$ 轴没有一个点代表 $a$,或在另一条轴上没有一个点可以说是 $b$,则这样的极限式应如何理解?这种情况希腊人早就发现了.$x$ 轴上有一个点(距原点为单位边长正方形的对角线长的那一点)不能用一个数(希腊人讲数实际上是有理数)来表示,因此 $x$ 趋向这样的点,就不能说成是 $x\to a$.总之,我们的极限概念实际上有一个默认的基础:直线上的点与实数之间有一个一一对应.从几何上看,这似乎是

不言而喻的，但是从逻辑上讲是有待证明的.实际上微积分基础中与这个没有明说的默认有关的许多基本的命题，如柯西序列必有极限，单调有界序列必有极限，其证明都有赖于此.

因此问题就在于如果只承认有理数，则作为数轴的直线上有一些点找不到对应的数，所以，有理数是有"空隙"的. 只有把这些空隙都填补起来，极限理论才有可靠的基础.我们不能采取这样的办法，即缺什么数就补什么数：例如，有理数中既然没有表示单位边长正方形的对角线长的数，就补进一个 $\sqrt{2}$，然后补上例如 $\sqrt[3]{7}$，e，π 等等.这样做何时才知道"所有的"空隙都补上了.(连什么叫"所有的空隙"都说不清楚.)因此我们需要一个系统的方法补上一些数，使得以后再也没有必要，没有可能再添补新数.怎样知道是"再也没有必要，没有可能"的呢？对此也必须作出数学上很明确的回答.

在进行这项工作前，我们要弄清我们现在所处的地位以及我们想要做到的事.

我们所处的地位是：我们已经有了有理数系 **Q**，它有以下的性质：

Ⅰ. **Q** 是一个域，即在其中可以进行加、乘及其逆运算减、除(但不允许以 0 为分母)等四则运算，而且加法与乘法均适合交换律与结合律，二者在一起则适合分配律.

Ⅱ. **Q** 是一个有序域，即对任意 $x,y\in\mathbf{Q}$，以下的三种关系必有一种且只有一种成立：$x<y$，$x=y$ 与 $x>y$.

这种次序关系有以下几个性质：

($Ⅱ_1$) 若 $x\leqslant y$，且 $y\leqslant x$，则 $x=y$；

($Ⅱ_2$) 若 $x\leqslant y$，$y\leqslant z$，则 $x\leqslant z$；

($Ⅱ_3$) 若：$x_1\leqslant x_2$，$y_1\leqslant y_2$，则 $x_1+x_2\leqslant y_1+y_2$；

($Ⅱ_4$) 若 $x\leqslant y$ 而 $z>0$，则 $xz\leqslant yz$.

Ⅲ. **Q** 是一个阿基米德域，即满足以下的阿基米德公理：任给 $x$、$y$ 而 $x>0$，$y>0$，则必一个正整数 $n$，使 $y\leqslant nx$

把Ⅰ—Ⅲ当作公理来看，这里面其实是有问题的，例如对自然数确有一组人们公认理——佩亚诺(G. Peano，1858—1932)公理而它并不就是Ⅰ.所以 **O** 中的算术运算的许多

柯西序列没有极限这种空隙要补上.这个补空的过程就称为完备化.我们不仅要问,完备化是否可能,还要问完备化的方式是否在某种意义下是唯一的?这个问题很重要.因此如果有不同的方法作完备化,而且作出了不同的实数系,则应该有不同的极限理论,有不同的微积分学.一旦出现这种情况,后果将是灾难性的.

2. **Q** 再补上一些数——称为无理数,说是"一些数"实在把完备化的内容说得太简单了,因为补上的新数要比原有的有理数多得多:原来的有理数只有可数无穷多个,而无理数是不可数无穷多个.这样构成了实数系 **R**.但是尽管补上的新数多得多,原有的 **Q** 却在 **R** 中稠密.正是这个稠密性保证了新数仍然适合上面对 **Q** 的要求 Ⅰ 和 Ⅱ.因此我们确实可以承认它们都是数.但是 Ⅲ 却还是在 **R** 上成立.因此,无穷小量与无穷大量都不是现在我们所理解的数,而它们是某种变量.

3. 有了实数系 **R**,而且得知它已经是完备的,我们就要用它来证明微积分的一些基本的命题,例如单调有界数列有极限存在.我们下面会讲到,在有了完备性以后,不但不会因柯西序列没有极限而发现原有的数系有需要填补的空隙,而且也不会例如有界单调数列没有极限而产生空隙.最终我们会看到,在一定的要求之下,也不会因为任何问题再出现空隙.就是说,我们对问题的解决是一劳永逸的.

现在我们就按这个计划来建立实数系.

第一步是介绍完备化过程.

我们的基本思想是:柯西序列就是实数,这本来是很直观的,例如,凡讲到 $\pi$,人们心目中就一定是指一个序列:

$$\{a_m\}=\{3,3.1,3.14,3.141,3.1415,3.14159,\cdots\}.$$

即 $\pi$ 的无穷小数展开式的前有限项.它显然是一个柯西序列,因为例如其中第 $m$ 项 $a_m$ 是 $\pi$ 的展开式到小数点后 $m-1$ 位,第 $n$ 项 $a_n$ 则是展开式小数点后前 $n-1$ 位.设 $m<n$,则 $|a_m-a_n|$ 小数点后前 $m-1$ 位全都是 0,因而 $|a_m-a_n|\leqslant\frac{1}{10^{m-1}}$,所以当 $m,n$ 充分大时,$|a_m-a_n|<\varepsilon$,$\varepsilon>0$ 是任意指定的有理数.所以它是柯西序列.为什么要说 $\varepsilon$ 是有理数呢?因为现在只有有理数才有定义.所以现在有理数的柯西序列定义应是 $\{a_n\}$,$a_n\in\mathbf{Q}$,对任意正有理数 $\varepsilon>0$,必可找到正整数 $N(\varepsilon)$,使当 $m,n>N(\varepsilon)$ 时 $|a_m-a_n|<\varepsilon$.但在以后定义了实数以后,柯西序列中的 $\varepsilon>0$ 就可以改说是"任意正的实数 $\varepsilon>0$",不过读者会看到即令不作这样的更改也没有影响.所以下面我们就不再强调这一点了.我们想要强调的是,$\pi$ 还可以表示为另一个柯西序列.

$$\{b_n\}=\{4,3.2,3.15,3.142,3.1416,3.14160,\cdots\}$$

所以,说一个柯西序列定义一个实数,并不恰当.不过 $a_n-b_n\to0$(读者可能会问,不是现在还没有定义极限吗?为什么能使用极限的记号呢?其实说极限没有定义是指序列、极限、$\varepsilon$,只要可能不是有理数时都没有定义.现在我们的 $\{a_n\}$,$\{b_n\}$,$\varepsilon$,… 都是有理数,这时甚至使用极限的 $\varepsilon$—$N$ 定义都没有困难).因此若有两个柯西序列 $\{a_n\}$、$\{b_n\}$,适合 $a_n-b_n\to0$,就说它们是等价的,记作

$$\{a_n\}\sim\{b_n\}\ (\text{即指 } a_n-b_n\to0).\tag{1}$$

**定义 1**　有理数的柯西序列按(1)式的等价类就是一个实数.

有的读者可能读过通常的实数理论，那里是以戴德金分割作为实数之定义的.何以这里要标新立异？何况现在我们的作法反而麻烦一些？理由在于今后我们将仿照由有理数构造实数的方法来构造许多函数空间，特别是 $L^p$ 空间.那时戴德金分割就用不上了.

于是，这个等价类中每一个具体的柯西序列都是这个等价类的代表元.在一个等价类中取哪一个柯西序列为代表元都是一样的.

有理数也是实数，那么哪一些等价类代表有理数呢？其实，若 $r$ 是一个有理数，令 $a_n=r$，得 $\{r,r,\cdots,r,\cdots\}$（记作 $\{r\}$）也是一个柯西序列，我们不妨称为常驻序列.我们不妨把 $r$ 本身与 $r$ 构成的常驻序列集合一一对应，因此，若一个等价类含有一个常驻序列 $\{r\}$，我们就认为这个等价类代表有理数 $r$.这里要注意，若一个等价类中含有常驻序列，则它不能含两个不同的常驻序列 $\{r\}$ 与 $\{s\}$，$r\neq s$.事实上若 $r-s\to 0$ 则必有 $r=s$ 而与 $r\neq s$ 矛盾，所以 $\{r\}$ 与 $\{s\}$ 不可能在同一个等价类中.于是我们把柯西序列等价类的集合分为两类：Ⅰ是含有某个常驻序列 $\{r\}$ 的等价类之集合，Ⅱ类之元是不含常驻序列的等价类.

**定义 2** Ⅰ类之元称为有理数，$\{r\}$ 对应于 $r$，Ⅱ类之元称为无理数.两类合并即得实数系 $\mathbf{R}$.

本来现在应该就来说明 $\mathbf{R}$ 为什么就是完备的，但在这以前，我们先来讨论实数的运算和次序.

第二步，实数的运算和次序关系.

实数的运算是比较容易定义的，设有两个实数 $a$ 与 $b$，并取其代表元 $\{a_n\}$ 和 $\{b_n\}$，则例如 $a+b$ 即定义为 $\{a_n+b_n\}$ 之等价类；$ab=\{a_nb_n\}$.我们要注意，$\{a_n+b_n\}$，$\{a_nb_n\}$ 仍是有理数的柯西序列，所以这些运算是有定义的.加法零元 0 是包含常驻序列 $\{0,0,\cdots,0,\cdots\}$ 的等价类；乘法单位元则是包含常驻序列 $\{1,1,\cdots,1,\cdots\}$ 的等价类.$\dfrac{1}{a}$ 是 $\left\{\dfrac{1}{a_n}\right\}$ 的等价类，不过表面上应要求 $a_n\neq 0$，而不仅是 $\{a_n\}\not\sim\{0,0,\cdots,0,\cdots\}$，但实际上，如果 $a\neq 0$，则 $\{a_n\}$ 中只会含有最多有限多个 $a_k=0$.否则，若有无限多个 $a_{n_k}=0$，$k=1,2,\cdots$，$n_k\to+\infty$，则对任一 $\varepsilon$，因为当 $m,n>N(\varepsilon)$ 时 $|a_m-a_n|<\varepsilon$，我们只要取 $n=n_k>N$ 则 $|a_m|=|a_m-a_{n_k}|<\varepsilon$，当 $m>N(\varepsilon)$ 时成立，即是说，$a=\{a_n\}\sim 0$，这与 $a\neq 0$ 矛盾.既然在 $a_n$ 中只有有限多个为 0，可以从 $\{a_n\}$"砍掉"前面有限多项，余下来仍然是一个柯西序列，仍然与原来的 $\{a_n\}$ 等价，从而仍是 $a$ 的代表元.但现在一切 $a_n\neq 0$，所以非零的 $a$ 一定有一个代表元 $\{a_n\}$ 其一切项均不为 0：$a_n\neq 0$.这样，$\dfrac{1}{a}=\left\{\dfrac{1}{a_n}\right\}$ 仍为柯西序列就很清楚了.总之，我们构造的新的数系：有理数柯西序列之等价类的集合，亦即实数系 $\mathbf{R}$ 仍是一个域（代数意义下的域）.

关于 $\mathbf{R}$ 是一个有序域的证明麻烦一些：我们可以很直观地想到，可以用 $a-b$ 为一正实数来表示 $a>b$，但为此对什么是正实数要认真论证一下. 而且我们可以更广泛一些，分别讨论什么是 $a=0$，$a>0$ 和 $a<0$，这里的 $a$ 表示某柯西序列 $\{a_n\}$ 的等价类，而 $a_n$ 则是有理数.$a=0$ 就是适合 $\lim\limits_{n\to\infty}a_n=0$ 的序列，这里我们又借用了有理数范围内的极限语言，而如前面所说这是许可的.如果不如此我们就要说：对任意 $\varepsilon>0$，必有 $N(\varepsilon)$ 存在而当 $n>N(\varepsilon)$ 时，$|a_n|<\varepsilon$.$a>0$ 问题复杂一些.我们决不能说 $a>0$ 就是 $\{a_n\}$ 中一切 $a_n>0$.因为例如 $\left\{\dfrac{1}{n}\right\}\sim 0$ 但 $\dfrac{1}{n}>0$.这里我们要利

用一个重要的事实，一个有理数 $s>0$ 时一定可以找到另一个 $r$ 使 $s>r>0$. 可否把这个事实用于 $\{a_n\}$？我们说：$a=\{a_n\}>0$ 就是存在一个正有理数 $r=\{r\}$ 以及一个正整数 $N$，使当 $n>N$ 时，$a_n>r>0$. 同样 $a=\{a_n\}<0$ 就是存在一个负有理数 $s=\{s\}$，以及一个正整数 $M$，使当 $n>M$ 时，$a_n<s<0$. 这样的定义行不行，就看是否每一个实数 $a$（设用有理数的柯西序列 $\{a_n\}$ 来表示）都能无疑义地规定其符号. 这是肯定的，因为 $\{a_n\}$ 是柯西序列，故对每个 $\varepsilon>0$ 必有 $N(\varepsilon)$ 存在，使当 $m,n\geqslant N(\varepsilon)$ 时，均有 $|a_m-a_n|<\varepsilon$. 所以对一切 $n>N(\varepsilon)$，

$$a_n=a_N+(a_n-a_N),$$

$$a_N-|a_n-a_N|\leqslant a_n\leqslant a_N+|a_n-a_N|.$$

这样 $a_n$ 之符号将与 $a_N$ 的大小有密切关系，这样就出现两个可能性. 其一是对于某一个 $\varepsilon>0$，$|a_N|>\varepsilon$. 如果 $a_N$ 为正，则 $a_N>\varepsilon$，由左边的不等式看 $a_n>a_N-\varepsilon=r>0$. 这样的 $\{a_N\}$ 符合前述 $a>0$ 之定义. 如果 $a_N$ 为负，则 $a_N<-\varepsilon$，由右边的不等式看 $a_n\leqslant a_N+\varepsilon=s<0$，这样的 $\{a_n\}$ 符合前述 $a<0$ 的定义. 第二个可能性是对一切 $\varepsilon$，按上法求出的 $a_N$ 均适合 $|a_N|\leqslant\varepsilon$. 这时当 $n>N(\varepsilon)$ 时，

$$|a_n|\leqslant|a_N|+|a_n-a_N|<2\varepsilon.$$

所以符合前述 $a=0$ 的定义. 总之任一实数 $a$ 在 $a>0$，$a=0$，$a<0$ 三种情况中必居其一且仅居其一. 实数 $a$ 既已可定义其符号，当然也就可以定义 $|a|$. 关于 $\mathbf{Q}$ 为有序域的四条性质（$\text{II}_1$）～（$\text{II}_4$）可以证明，对 $\mathbf{R}$ 也都适用. 但是证明却可能比较冗长.

阿基米德公理对于 $\mathbf{R}$ 仍然是成立的. 这一点也不来证明了. 它的重要性下面就会看到.

第三步是关于实数系 $\mathbf{R}$ 的基本性质的证明. 现在我们已经有了实数系了. 此时不妨回顾一下，什么是实数？实数就是一个可以对之进行四则运算，并可比较大小，且适合阿基米德公理的对象. 其实，我们对数的领悟已经有了一个飞跃. 比方有理数，朴素的理解就是一个屈指可数的数. 它的大小，一则可以用扳手指头的方法来比较（用抽象的数学语言来说，就是与“手指头的集合”一一对应）. 也可以通过几何方法用“量”来比较：圆内接多边形因为只是圆的一部分，部分小于整体，所以这些内接多边形面积小于圆面积（《几何原理》上就是这样讲的）. 但是到了实数就一般地没有这种朴素的理解了. 其四则运算化成了有理数柯西序列的运算. 其大小则归结为与0的比较，即正、负号. 读者们会问，实数诚然可以比较大小，但“它自己”有没有“大小”？上面说的思想上的飞跃在此：能比较大小就是有大小. 所以圆面积不必看成内接正多边形面积 $P_n$（其实 $P_n$ 一般已不是有理数）序列的极限，而是“序列”本身. 这一些读者们应该都懂得了. 但是我们的直觉还能不能保存？说算术化就是完全抛弃几何直观，至少在心理上是做不到的. 例如说，$\lim P_n=\pi r^2$ 至少在心理上意味着内接正多边形边数越多，$P_n$ 越接近圆的面积，这一点是无法排除的. 因此，说实数 $a$ 由有理数的柯西序列 $\{a_n\}$ 决定，在我们的直观上仍旧是 $\lim\limits_{n\to\infty}a_n=a$. 算术化以后能否保持这个直觉的理解呢？首先要在实数系 $\mathbf{R}$ 中把 $\lim\limits_{n\to\infty}a_n=a$ 讲清楚. 这就是任给一个实数 $\varepsilon>0$（不再是有理数 $\varepsilon>0$ 了），必可找到正整数 $N$，使当 $n>N$ 时，$a_n$ 与 $a$ 之距离小于 $\varepsilon$. 可是 $a_n$ 是有理数，$a$ 是实数，怎样讨论它们的距离呢？这就需对有理数有另一个理解：有理数就是常驻序列 $\{a_n,a_n,\cdots,a_n,\cdots\}$. 于是我们有一个一般的柯西序列 $a=\{a_1,a_2,\cdots,a_n,\cdots\}$，一个常驻序列 $\{a_n,a_n,\cdots,a$

列$\{\varepsilon,\varepsilon,\cdots,\varepsilon,\cdots\}$)比较大小. 这样就会有

**基本定理Ⅰ*** 任一个有理数柯西序列$\{a_1,\cdots,a_n,\cdots\}$必收敛于它所定义的实数.

**证** $a_n-a=\{a_n-a_1,a_n-a_2,\cdots,a_n-a_m,\cdots\}$. 这里 $n$ 是固定的,而 $m$ 则取一切正整数值. 由于$\{a_1,a_2,\cdots,a_m,\cdots\}$是柯西序列,对任意 $\varepsilon$ 必可找到$N(\varepsilon)$,使当 $m,n>N(\varepsilon)$时,$|a_n-a_m|<\varepsilon$. 现在在 $a_n-a$ 的序列表示中,$n>N(\varepsilon)$已是固定的,而当 $m>N(\varepsilon)$时,又有

$$-\frac{\varepsilon}{2}<a_n-a_m<\frac{\varepsilon}{2},$$

所以

$$\varepsilon-(a_n-a_m)>\varepsilon-\frac{\varepsilon}{2}=\frac{\varepsilon}{2},$$

$$\varepsilon+(a_n-a_m)>\varepsilon-\frac{\varepsilon}{2}=\frac{\varepsilon}{2}.$$

作为定义实数的柯西序列来看,就是

$$\{\varepsilon,\cdots,\varepsilon,\cdots\}-(a_n-a)>0,$$
$$\{\varepsilon,\cdots,\varepsilon,\cdots\}+(a_n-a)>0.$$

亦即

$$\varepsilon>a_n-a,$$
$$-\varepsilon<a_n-a,$$

或者说,当 $n>N(\varepsilon)$时

$$|a_n-a|<\varepsilon.$$

定理证毕.

**推论1** 有理数系 $\mathbf{Q}$ 在实数系中处处稠密.

**证** 所谓处处稠密就是指:任给一个实数 $a$,必可找到一个有理数列$\{a_n\}$,使$\lim\limits_{n\to\infty}a_n=a$. 但上面证明的就是这件事.

这里附带提醒一下,当我们说"找到一个有理数序列$\{a_n\}$"时,这里的 $a_n$ 可以按朴素的有理分数 $q/p$ 的意义来理解,而说$\lim\limits_{n\to\infty}a_n=a$ 时,$a_n$ 一定是指常驻序列$\{a_n,a_n,\cdots,a_n,\cdots\}$的等价类.

这个推论的重要性在于:它告诉我们,$\mathbf{Q}$ 中"空隙"太多了,比所有有理数还多,因为有理数只是可数无穷多个;尽管如此,仍然处处有有理数. 现在的问题是用以上方式是否又会产生新的空隙?因此是否仍有必要再令实数组成的某些柯西序列为另一类新数,把 $\mathbf{R}$ 再扩充一些呢?下面的基本定理表明不再会有这样的情况:

**基本定理Ⅰ** 任一个实数组成的柯西序列$\{a_n\}$必定义一个实数 $a$,使得$\lim\limits_{n\to\infty}a_n=a$.

**证** 由推论1,对$\{a_n\}$中的每一个实数 $a_n$ 都可以找到一个有理数$a_n'$,使$|a_n-a_n'|<\dfrac{1}{n}$,这样一来,$|a_n'-a_m'|\leqslant|a_n'-a_n|+|a_n-a_m|+|a_m-a_m'|<\dfrac{1}{n}+\dfrac{1}{m}+|a_n-a_m|$,对任一 $\varepsilon>0$,由$\{a_n\}$为柯西序列可知,一定有 $N(\varepsilon)$在,当 $n,m>N(\varepsilon)$时$|a_m-a_n|<\dfrac{\varepsilon}{3}$,我们还可以取$N(\varepsilon)>\dfrac{3}{\varepsilon}$,于是当 $n,m>N(\varepsilon)$时还有$\dfrac{1}{n},\dfrac{1}{m}<\dfrac{\varepsilon}{3}$,因此当 $n,m>N(\varepsilon)$时

$$|a_n' - a_m'| < \frac{1}{n} + \frac{1}{m} + |a_n - a_m| < \varepsilon.$$

因此,$\{a_n'\}$是一个有理数的柯西序列,由基本定理 Ⅰ* 及推论 1 即有

$$\lim_{n\to\infty}(a_n - a) = \lim_{n\to\infty}(a_n - a_n') + \lim_{n\to\infty}(a_n' - a) = 0.$$

基本定理 Ⅰ 告诉我们,如果把没有极限的柯西序列称为“空隙”,则经过以上的完备化手续以后,实数系 **R** 中再也没有“空隙”了,因此是完备的. 而且有理数系 $\mathbf{Q}\subset\mathbf{R}$ 在 **R** 中稠密. 有没有在其它意义下产生的空隙呢?以下我们会证明,在一定意义下再也不会有空隙了,因此 **R** 确实是完备的.

**2. 完备性和实数系的一些基本定理** 微积分中有一些基本定理,其实都是想证明有一个具某种性质的实数存在. 如果证明不了,我们就会想,能不能再用上面的方法来定义新数来扩充 **R** 呢?如果这样,就可以说 **R** 中仍有空隙. 这些定理的证法,最后都是归结为寻求一个柯西序列,并利用基本定理 Ⅰ 得知有一个实数 $a$ 存在,然后再证明 $a$ 就是所寻求的有某种性质的实数. 这样的定理我们可以举出三个来,并分别利用基本定理 Ⅰ 加以证明.

**定理 2**(**区间套定理**) 设有实数闭区间套,即闭区间之序列$\{[a_n, b_n]\}$,而且适合$[a_n, b_n]\subset[a_{n-1}, b_{n-1}]$. 若这些区间之长$(b_n - a_n)\to 0$,则必存在唯一实数 $c$,含于所有区间$[a_n, b_n]$之中.

**证** 任给一个 $\varepsilon>0$,必有正整数 $N(\varepsilon)$存在,使当 $n>N(\varepsilon)$时,$b_n - a_n<\varepsilon$,现在固定这样一个 $m>N(\varepsilon)$,则 $a_{m+1}, a_{m+2}\cdots$均含于$[a_m, b_m]$中,因此$|a_{m+k} - a_m|\leqslant|b_m - a_m|<\varepsilon$,这样$\{a_n\}$形成一个实数的柯西序列,因而由基本定理 Ⅰ,$\lim\limits_{n\to\infty} a_n = c$ 存在. 很容易证明 $a_m\leqslant c\leqslant b_m$,即 $c\in[a_m, b_m]$. 而由 $m$ 可以任意选取,只要 $m>N(\varepsilon)$即可,选 $m+k$ 代替 $m$ 易见 $c\in[a_{m+k}, b_{\cdot}$
$k=1,2,\cdots$,又由于$[a_1, b_1]\supset\cdots\supset[a_{m-1}, b_{m-1}]\supset[a_m, b_m]$所以 $c$ 也在$[a_1, b_1],\cdots,[$
$b_{m-1}]$中,总之 $c$ 含于一切$[a_n, b_n]$中. 这样的 $c$ 必是唯一的,因为设有另一个 $c'\in[a_n, b_n]$
$1,2,\cdots$,则因 $c'\neq c$,$|c'-c|$为一确定的正数,而且$|c'-c|\leqslant|b_n - a_n|$对一切 $n$ 均成
$b_n - a_n\to 0$,这与 $c'\neq c$ 矛盾.

**定理 3**(**单调有界序列收敛定理**) 设$\{a_n\}$是单调上升的有界序列:$a_1\leqslant a_2\leqslant\cdots\leqslant M$.则

所以由上式

$$a_{n_k'} \geqslant k\varepsilon_0 + a_{n_1}.$$

令 $k \to +\infty$,即有 $a_{n_k'} \to +\infty$ 而与 $\{a_n\}$ 是有界序列矛盾.由基本定理Ⅰ有

$$\lim_{n\to\infty} a_n = c.$$

定理证毕.

当然,对于有界的单调下降序列,本定理也成立.

**定理 4** 有界集必有确界.

**证** 我们只证有界集 $A$ 必有下确界 $\inf A = m$. 为方便计,用十进小数来表示 $A$ 的下界,于是先设 $m_0$ 是整数,$m_0$ 是 $A$ 的下界,但 $m_0+1$ 不是下界.取 $m_1$ 为 0 到 9 间的某个数字,使 $m_0 + m_1/10$ 是 $A$ 的下界,但是 $m_0 + (m_1+1)/10$ 不是下界.确定了 $m_1$ 以后再看小数点后的下一位.再取一个整数 $0 \leqslant m_2 \leqslant 9$,使 $m_0 + \frac{m_1}{10} + \frac{m_2}{10^2}$ 是 $A$ 的下界,但 $m_0 + \frac{m_1}{10} + \frac{(m_2+1)}{10^2}$ 不是下界.仿此,可以得到一个无尽十进小数 $m_0.m_1m_2\cdots m_k\cdots$(若从某个 $k$ 起 $m_{k+1} = m_{k+2} = \cdots = 0$,我们实际上得到的是一个有尽小数.但是我们仍然视这为无尽小数).它的任何"一段"$m_0.m_1\cdots m_k$ 都是下界,而 $m_0.m_1\cdots(m_k+1)$ 不是下界(不过如果 $m_k = 9$,$m_k+1 = 10$,就表示小数中要进一位).但是无尽十进小数必是一个柯西序列.因此由基本定理Ⅰ,它表示一个实数 $c$,今证 $c$ 是 $A$ 的下确界.

任取 $A$ 中一元 $a \in A$,由于 $m_0.m_1\cdots m_k$ 是下界,所以

$$m_0.m_1\cdots m_k \leqslant a.$$

令 $k \to +\infty$ 即有

$$c \leqslant a.$$

所以 $c$ 是 $A$ 的下界.为了证明它是下确界,我们再取任意数 $d > c$,今证 $d$ 必不是下界.因为 $d$ 的十进小数 $n_0.n_1\cdots n_k\cdots$ 中至少有一位大于 $c$ 的十进小数的相应数字,设 $n_i$ 中第一个大于 $m_i$ 的是 $n_k$,即 $n_k \geqslant m_k + 1$,所以

$$d \geqslant n_0.n_1\cdots n_k \geqslant m_0.m_1\cdots(m_k+1).$$

但是右边必非 $A$ 之下界,所以 $d$ 也不是 $A$ 的下界,于是 $c = \inf A$.

同样可以证明上有界集必有上确界存在.

还有一些重要的定理,因为它们是关于实数空间 $\mathbf{R}$ 的另一重要问题——紧性问题的,我们将在下面再讲.我们现在还是把涉及完备性的话说完.

定理 2~4 与基本定理Ⅰ是等价的:即是说,不但可以由基本定理Ⅰ证明定理 2~4,其实由定理 2~4 的任何一个也可以证明基本定理Ⅰ和另两个定理.其原因是,这些结果都是以不同方式回答了一个问题:有理数系 $\mathbf{Q}$ 的空隙应该怎样去刻画并填补起来.例如区间套定理讲的就是:一个有理数区间套若其各区间长度趋向 0,可能会出现在有理数系 $\mathbf{Q}$ 中找不到一个公共元,这样就出现了空隙,而定理 3 告诉我们:没有关系,有了这样的区间套,就有一个柯西序列例如 $\{a_n\}$,它定义一个新数(实数)$a$,把这个 $a$ 填到空隙里去,这个区间套就有了唯一的公共点 $a$,而区间套定理就得到了证明.很自然,我们也可以走另一条路,如果某个有理数区间套出了空隙——即

不能以某有理数为公共点，就说它定义了一个新数，称为实数. 把这个实数"塞"进区间套的每个区间内，于是此区间套就有了一个公共点（即此实数）. 这样定义了的实数系仍可进行四则运算，仍有大小（还可加上阿基米德公理），于是此区间套可以说是趋向这个实数的，而原来的有理数系按原有的大小说也好，按扩充成的实数系的大小来说也好都在此实数系中稠密（相应于基本定理Ⅰ* 和推论 1），如果再用扩充成的实数系作区间套，就不会再出现空隙，而知它必含有唯一公共点（相应于基本定理Ⅰ，即定理 2）. 把它称为基本定理Ⅲ，则由它不但可以证明定理 3，4，还可以证明关于极限存在的柯西准则：任一实数柯西序列必收敛于唯一实数. 总之这几个定理都可以作为基本定理，但它的证明必须分成两步：先是基本定理 $X^*$：这是由有理数系 $\mathbf{Q}$ 填补空隙而得实数系 $\mathbf{R}$；然后再是基本定理 $X$，这是说实数系 $\mathbf{R}$ 中再没有空隙了，这就是说，上面的基本定理和三个定理实际上给出了构造实数系 $\mathbf{R}$ 的等价的方法.

于是还有一个问题：会不会在某种新环境下 $\mathbf{R}$ 出现新的空隙，需要用新的方法填补，而得到一个与上面作出的 $\mathbf{R}$ 不同的实数系 $\widetilde{\mathbf{R}}$ 呢？例如在通常较详细的微积分教本上都是用的戴德金分割来定义实数系 $\mathbf{R}$ 而与以上诸定理均不同. 我们甚至应想到，还会不会有其它方法作出 $\mathbf{Q}$ 的不同扩充来呢？下面的定理告诉我们：不会.

**基本定理Ⅱ**　有理数系 $\mathbf{Q}$ 扩充为完备的实数系并使 $\mathbf{Q}$ 在其中稠密的方法在等距同构意义下是唯一的.

**证**　先把定理的含意解释清楚. 设我们已把有理数系 $\mathbf{Q}$ 扩充为实数系 $\widetilde{\mathbf{R}}$，而且使 $\widetilde{\mathbf{R}}$ 中的运算关系与次序（即大小）关系与 $\mathbf{Q}$ 中原有的运算关系次序关系相协调；例如设 $a\in\mathbf{Q}\subset\widetilde{\mathbf{R}}$，则 $a$ 在 $\mathbf{Q}$ 中的大小（即原来的大小）与它在 $\widetilde{\mathbf{R}}$ 中的大小一致. $\mathbf{Q}$ 在 $\widetilde{\mathbf{R}}$ 中稠密，即 $\widetilde{\mathbf{R}}$ 中任一元 $r$ 必可用 $\mathbf{Q}$ 中的元之序列 $\{a_n\}$ 去逼近：$|a_n-r|\to 0$. 这里的距离既可说是 $\mathbf{Q}$ 中的距离，也可说是 $\widetilde{\mathbf{R}}$ 中的距离. 这个定理说，如果有两种适合这样的要求的扩充 $\widetilde{\mathbf{R}}_1$ 与 $\widetilde{\mathbf{R}}_2$，则二者必等距同构，即是说 $\widetilde{\mathbf{R}}_1$ 中的"实数"与 $\widetilde{\mathbf{R}}_2$ 中的"实数"，无论就其大小而言或者就其运算关系而言都是一样的. 但是实数除了是运算对象与大小关系而外还有什么含义呢？所以我们说 $\widetilde{\mathbf{R}}_1$ 和 $\widetilde{\mathbf{R}}_2$ 是一样的. 因此，$\mathbf{Q}$ 的扩充是唯一的. 明白了定理的含意后，证明是很容易的. 我们记上面讲的用柯西序列定义作扩充的实数系为 $\mathbf{R}$. 今证任何适合定理要求的扩充 $\widetilde{\mathbf{R}}$ 必与 $\mathbf{R}$ 等距同构.

设有 $\tilde{r}\in\widetilde{\mathbf{R}}$，则由 $\mathbf{Q}$ 在 $\widetilde{\mathbf{R}}$ 中稠密，必可找到 $\mathbf{Q}$ 中一个序列 $\{a_n\}$，$a_n\to\tilde{r}$，这个 $\{a_n$

重要性的数学家大概是博雷尔(E.Borel),他是从这样一个事实开始的.

**定理 5(海涅-博雷尔(Heine-Borel)定理)** 设$[a,b]$是$\mathbf{R}$中的一个闭区间,$\{U_\lambda\}$是它
个开覆盖,则从$\{U_\lambda\}$中一定可以取出有限多个$U_1,\cdots,U_k$,构成$[a,b]$的一个子覆盖:

$$[a,b]\subset\bigcup_{j=1}^{k}U_j.$$

**证** 在证明此定理之前,我们还是先来解释一下它的意义."覆盖"的含意很简单,这里
说$[a,b]\subset\bigcup_\lambda U_\lambda$.需要注意的是,$\{U_\lambda\}$不一定是可数多个$U_\lambda$(若$U_\lambda$为可数多个叫可数覆
时我们需取$\lambda=1,2,\cdots$),所以$\lambda$就只是一个参量,它的集合$\Lambda$可以是有限集,也可以是可
或不可数集.但一定不是空集,因为那样一来,就没有$U_\lambda$了,也就不会覆盖什么.开覆盖
$\{U_\lambda\}$之每个元都是开集.但我们知道,在$\mathbf{R}$中每个开集都是最多可数多个开区间之并,所以
组成各个$U_\lambda$的开区间分开来重新编号,并仍记作$\{U_\lambda\}$,则每个$U_\lambda$又都可以设为开区间.
因为这个定理太重要了,所涉及的概念不只会被应用在$\mathbf{R}$上,那时也没有开区间的概念.所
们现在还是说$U_\lambda$就是一个开集.

但是我们的证明中还是要利用到每个$U_\lambda$均可分为若干个开区间.其证明其实很容易

若$[a,b]$只需有限多个$U_\lambda$即可覆盖,则没有什么好证的了.若不然,将它平分为两个
间:$[a,b]=\left[a,\frac{a+b}{2}\right]\cup\left[\frac{a+b}{2},b\right]$(两个小闭区间有一个公共点$\frac{a+b}{2}$并无关系),则至少
个需无限多个$U_\lambda$才能覆盖.设这一个是$[a_1,b_1]$,再把$[a_1,b_1]$平分为二,又至少有一个需
多个$U_\lambda$才能覆盖.仿此,我们会得到一个区间套$[a_k,b_k]$,其长为$\frac{1}{2^k}(b-a)\to 0$,由区间套
必定有一点$c$在所有$[a_k,b_k]$中.因为$\{U_\lambda\}$覆盖了$[a,b]$,所以$c$必在某个$U_{\lambda_0}$之内,从而必
一开区间(不妨就设$U_{\lambda_0}$是一个开区间)内,由开区间的性质,必有以$c$为中点的小开区间,(
$c+r)\subset U_{\lambda_0}$,其长为$2r$.但$c$又在一切$[a_k,b_k]$中,若某个$[a_k,b_k]$之长小于$r$,必有$[a_k,$

由 只要 充分大 [ ]之长自然小于 所以上述平分只要进行 步,则根本无需

$\bigcup_k U_k$ 是明显的事，但是$\{U_k\}$中决没有 **R** 的有限子覆盖.因为若要覆盖 **R**，就需覆盖其上的无穷多个坐标为整数的点.但每一个区间 $U_k$ 只可能含一个“整点”$k$，因此要覆盖 **R** 就必须用完全部的 $U_k$，而不可能有有限子覆盖. **R** 整个说来不紧，但每一点都有一个邻域(不妨设为一区间$(a,b)$)，其闭包$[a,b]$为紧.这种情况称为局部紧.所以 **R** 不紧而仅只为局部紧.紧与局部紧是有重大区别的，这种区别在许多数学问题上有重大影响.

**R** 中与紧性有关还有几个概念，一是魏尔斯特拉斯-波尔查诺性质.一个空间 $K$ 有此性质是指此空间中任一序列$\{x_n\}\subset K$，必有一个位于 $K$ 内的极限点，即这样的点 $x_0$：其任一邻域 $I=(x_0-\varepsilon,x_0+\varepsilon)$与 $K$ 之交 $I\cap K$ 中均有此序列中的点.另一是列紧性：我们说空间 $K$ 是列紧的即指$K$ 中之任一序列$\{x_n\}\subset K$ 必有一个收敛子序列，其极限 $x_0\in K$，这里我们有.

**定理6**　设 $K\subset$**R**，其中之开、闭集即指 **R** 中的一个开、闭集与 $K$ 之交，这时以下三个命题等价.

(i) $K$ 为紧.

(ii) $K$ 具有魏尔斯特拉斯－波尔查诺性质.

(iii) $K$ 具有列紧性.

这个定理的证明并不太难，而主要涉及一些技术性的细节，所以我们略去它.但重要的是要指出紧与列紧、与魏尔斯特拉斯－波尔查诺性质是不相同的概念.只不过对 **R** 的子集 $K$ 三者一致，而在更广泛的情况下就不一定一致了.若 $K=[a,b]$，则(ii)与(iii)都是微积分教程中常见的(但是重要的)结果.正因为如此，有些读者会不注意把它们区别开来.

微积分教本中通常都要讲到闭区间$[a,b]$上的连续函数的一些基本性质.我们已经知道$[a,b]$是 **R** 的一个紧集，而通常微积分教本中通常回避使用这个词，我们则坚持使用紧集这个词，除非是必须用到$[a,b]$作为一个区间才有的性质时，说明我们讨论的是区间. 例如下面的(iii). 其目的是向读者“灌输”紧性之重要. 我们将要证明下面的定理.

**定理7**　设 $K\subset$**R** 是一个紧集，$f(x)$是 $K$ 上的连续函数，则有

(i) $f(x)$在 $K$ 上有界，而且必能达到其上、下确界使之成为最大、最小值.

(ii) $f(x)$必在 $K$ 上一致连续.

(iii) 若 $K=[a,b]$，则 $f(x)$可以取到其最大、最小值之间的一切值.

**证**　在证明之前，先对定理作一些解释. 我们上面特别提醒，紧集 $K$ 不一定是一个区间，但讲函数 $f(x)$在一点 $x_0$ 连续时，$x_0$ 总是取为某一开区间的内点.现在所谓 $f(x)$在 $x_0\in K$ 连续，是指在$\lim\limits_{x\to x_0}f(x)=f(x_0)$中 $x$ 必须在$K$ 内趋向 $x$.这在通常的微积分教本中本已见过，例如说 $f(x)$在$[a,b]$的左端 $x_0=a$ 处连续，就要求 $x$ 从$a$ 点右方(即$[a,b]$内部)趋向 $a$ 点，并称为右连续，其实就是 $f(x)$相对$[a,b]$的连续性的意思.甚至 $K$ 还可以是孤立点(有限多个点所成的集合一定是紧的)$x_0$，这时$\lim\limits_{x\to x_0}f(x)=f(x_0)$中 $x\to x_0$ 只能解释为 $x=x_0$，所以 $f(x)$只要在 $x_0$ 处有定义就是 $f(x)$在 $x_0$连续. (同理，$\lim\limits_{x\to x_0}f(x)=A$ 就可能没有意义了，因为这里既要求 $x\in K$，又要求 $x\neq x_0$，二者可能矛盾.)与此类似.当我们说到 $K$ 的开覆盖$\{U_\lambda\}$时，所谓 $U_\lambda$ 是开集，是指 $U_\lambda$ 是 **R** 中一个开集(它是若干个开区间之并)$\widetilde{U}_\lambda$ 与$K$ 之交：$U_\lambda=\widetilde{U}_\lambda\cap K$.甚至为简单起

见我们就说 $U_\lambda$ 是 $(\alpha,\beta)$. 这种问题在本书中已提到过好几次.在第三节讲到拓扑空间的子空间时我们再系统地说明.下面开始证明.

(i) 因为 $f(x)$ 是连续的,故对任一点 $x_0\in K$,对某个固定的 $\varepsilon_0$ 必有 $\delta>0$ 存在使在 $U_{x_0}=(x_0-\delta,x_0+\delta)$ 内(准确些说是在 $U_{x_0}=(x_0-\delta,x_0+\delta)\cap K$ 内),有

$$f(x_0)-\varepsilon_0<f(x)<f(x_0)+\varepsilon_0. \tag{4}$$

于是 $\{U_{x_0}\}$(参数 $\lambda$ 现在就是 $x_0$,$\Lambda$ 就是 $K$)成为 $K$ 的一个开覆盖. 而因 $K$ 为紧,可以找到一个有限子覆盖 $\{U_i\}\quad i=1,2,\cdots,k$. 在 $U_i$ 中,

$$m_i=f(x_i)-\varepsilon_0<f(x)<f(x_i)+\varepsilon_0=M_i\quad i=1,2,\cdots,k. \tag{5}$$

令 $M=\max\limits_{1\leqslant i\leqslant k}M_i$, $m=\min\limits_{1\leqslant i\leqslant k}m_i$,则知在 $K$ 上

$$m\leqslant f(x)\leqslant M.$$

因此 $f(x)$ 有界.

既有界必有上下确界,令 $\mu$ 为其上确界,我们用反证法证明一定有一点 $x^*\in K$ 使 $f(x^*)=\mu$. 如果这样的 $x^*$ 不存在,则在 $K$ 上必有 $\mu-f(x)>0$. 令 $\varphi(x)=\dfrac{1}{\mu-f(x)}$,则 $\varphi(x)$ 在 $K$ 上连续,而如上所证,$\varphi(x)$ 在 $K$ 上必有上界 $\lambda$,$\dfrac{1}{\mu-f(x)}\leqslant\lambda$. 因此 $f(x)\leqslant\mu-\dfrac{1}{\lambda}<\mu$ 在 $K$ 上成立,因此 $\mu$ 不是上确界.命题得证.

我们要看一下紧性在这个证明中起的作用.关键在于(4)式本来对一切 $U(x_0)$ 成立,所以在任一 $U_{x_0}$ 中有 $m(x_0)\leqslant f(x)\leqslant M(x_0)$. 但只要有无穷多个 $x_0$ 点,就有可能有无穷多个不同的 $M(x_0)$,由其中不一定能找出一个最大的(有时为方便起见我们也说其最大是 $+\infty$),紧性则把无穷多个(4)归结为有限多个(5),而由有限多个 $M_i$ 中自然有最大的 $M$(而自然是指有限的实数)存在.所以紧性的作用在于能把无穷多个对象归结为有限多个对象.人们常说,紧集是“最接近于有限集”的无穷集合就是这个意思.

上面说有时为方便起见常说无上界的函数就是以 $+\infty$ 为上界的函数,而且这时上确界就是 $+\infty$. 按阿基米德公理,不可能有一个实数是无穷大,所以这只是一个方便的说法,但我们时常要用这种说法,本节之末我们再稍作解释.

(ii) 对任意 $\varepsilon>0$,对 $K$ 中每一点 $x^*$ 必有 $\delta^*(\varepsilon,x^*)$ 存在,使当 $x\in(x^*-\delta^*,x^*+\delta^*)\cap K$ 时,

$$|f(x)-f(x^*)|<\frac{\varepsilon}{2}.$$

在此区间中任取两点 $\bar{x}$,$\bar{\bar{x}}$ 则有

$$|f(\bar{x}_1)-f(\bar{\bar{x}}_2)|\leqslant|f(\bar{x})-f(x^*)|+|f(\bar{\bar{x}})-f(x^*)|<\varepsilon. \tag{6}$$

把区间 $(x^*-\delta^*,\ x^*+\delta^*)$ 缩小一半,并记

$$U_{x^*}=\left(x^*-\frac{\delta^*}{2},x^*+\frac{\delta^*}{2}\right)\cap K,\quad x^*\in K,$$

则 $\{U_{x^*}\}$ 仍是 $K$ 的开覆盖,所以可以从其中选出有限多个:

$$U_i=\left(x_i-\frac{\delta_i}{2},x_i+\frac{\delta_i}{2}\right)\cap K,i=1,2,\cdots,k,$$

仍是 $K$ 的开覆盖. 取

$$\delta=\min_{1\leqslant i\leqslant k}\frac{\delta_i}{2}>0.$$

并且证明,$K$ 中任意两点 $\bar{x},\bar{\bar{x}}$,只要 $|\bar{x}-\bar{\bar{x}}|<\delta$,则这两点一定位于同一个 $(x_i-\delta_i,x_i+\delta_i)$ 中. 事实上,既然 $\{U_i\}$ 是 $K$ 的覆盖,则 $\bar{\bar{x}}$ 必在某一个 $U_i$ 中,不妨设在 $U_1$ 中. 因此

$$|\bar{\bar{x}}-x_i|<\delta\leqslant\frac{\delta_1}{2}<\delta_1,$$

$\bar{x}$ 和 $\bar{\bar{x}}$ 同在 $(x_1-\delta_1,x_1+\delta_1)$ 中,而由(6)式

$$|f(\bar{x})-f(\bar{\bar{x}})|\leqslant|f(\bar{x})-f(x_1)|+|f(x_1)-f(\bar{\bar{x}})|<\varepsilon.$$

所以 $f(x)$ 在 $K$ 上一致连续.

(iii) 的证明全不相同. 它实际上与紧性无关,但我们还是先利用紧性来证明它. 首先我们已知 $f(x)$ 必在两点达到最大值 $M$ 与最小值 $m$: $f(\alpha)=M,f(\beta)=m$. 我们不妨设 $\alpha<\beta$,并且用 $[\alpha,\beta]$ 代替 $[a,b]$. 我们要证对任一 $\mu,M>\mu>m$,在 $[a,b]$ 中有一点 $c$ 使 $f(c)=\mu$. 为此,用 $\varphi(x)=f(x)-\mu$ 代替 $f(x)$,则有 $\varphi(\alpha)>0,\varphi(\beta)<0$,因此我们把待证明的命题改为:若 $f(x)$ 在 $[a,b]$ 的两个端点异号,则一定存在一点 $c\in[a,b]$,使 $f(c)=0$. 我们仍用反证法,设这样的 $c$ 不存在. 任取一点 $x^*\in[a,b]$,必有 $f(x^*)\neq0$. 因为 $f(x)$ 是连续的,故当 $\delta$ 充分小时,$f(x)$ 在 $U_{x^*}=(x^*-\delta,x^*+\delta)$ 中与 $f(x^*)$ 符号相同. $\{U_{x^*}\},x^*\in[a,b]$ 构成 $[a,b]$ 的一个开覆盖,于是它有一个有限的子覆盖 $U_1,U_2,\cdots,U_k$. 我们不妨将它从左到右排列,而且假设它们都是开区间,这样 $a\in U_1,b\in U_k$ 而且相邻的两个 $U_i$ 必定相交: $U_{i-1}\cap U_i=I_i$ 是一个开区间. 在每个 $U_i$ 上,$f(x)$ 都是不变号的,于是在 $U_1$ 中 $f(x)$ 与 $f(a)$ 同号,在 $I_2$ 中取一点 $x_2$,则 $x_2\in U_1$,从而 $f(x_2)$ 与 $f(a)$ 同号. 但是在 $U_2$ 中 $f(x)$ 也是不变号的,而且 $x_2\in U_2$,所以在 $U_2$ 中 $f(x)$ 也与 $f(a)$ 同号. 仿此,在所有的 $U_1,\cdots,U_k$ 中 $f(x)$ 均与 $f(a)$ 同号. 从而 $f(b)$ 与 $f(a)$ 同号,而这与假设矛盾,故命题得证.

这个命题与前两个有何不同?我们仔细分析一下.

首先,它利用了紧性,这样我们只要推理有限步即可得到结论. 这与前面说的紧集是最接近有限集的无限集合是一致的. 其次证明中利用 $f(x)$ 连续而且 $f(x^*)\neq0$,因此,只要 $\delta$ 充分小,$f(x)$ 在 $U_{x^*}$ 中不变号. 造成矛盾的不只是 $f(x)$ 不变号,更重要的是,$U_1,\cdots,U_k$ 两两相连,有公共点 $x_i\in U_{i-1}\cap U_i$,这样才会造成矛盾. 若不是讨论 $[a,b]$,而是一般的 $K$,这种两两相连就可能不出现,例如设 $K=[0,1]\cup[2,3]$,它是一个紧集,于是由 $\{U_{x^*}\}$ 中可以找到有限子覆盖 $U_1,\cdots,U_k$,很可能 $U_1\cup U_2\cup U_3$ 覆盖了 $[0,1]$,$U_4\cup\cdots\cup U_k$ 覆盖了 $[2,3]$,而 $U_3$ 与 $U_4$ 可能不相交. 所以我们利用了 $[a,b]$ 的另一个重要特性——连通性——连通性与完备性、紧性一样是非常重要的性质. 许多重要的数学结果(例如这个命题)来自连通性. 因篇幅的原因,我们这里不再多说,§3 将专门讨论有关连通性问题.

总之关于实数系 $\mathbf{R}$,我们现在知道了它是一个有序的域,它满足阿基米德公理,而且它具有完备性、局部紧性与连通性.

最后我们讲一下关于 $\pm\infty$ 的问题.阿基米德公理已经告诉我们,它们不是实数,但是为方便起见,有时我们还是把它们“当数看待”,并作如下规定:

$$(\pm\infty)+(\pm\infty)=\pm\infty;$$

$$a+(\pm\infty)=\pm\infty;$$

$$a\cdot(\pm\infty)=\pm\infty \quad 若\ a>0;$$

$$a\cdot(\pm\infty)=\mp\infty \quad 若\ a<0;$$

$$(\pm\infty)\cdot(\pm\infty)=+\infty,\ (\pm\infty)\cdot(\mp\infty)=-\infty;$$

$(\pm\infty)-(\pm\infty)$ 无定义,$0\cdot(\pm\infty)$ 有的书上说是无定义,有的书上则规定 $0\cdot(\pm\infty)=0$.

引进 $\pm\infty$,并且得到所谓“扩充的实数系”,纯粹只是为了讲话方便,例如上无界集就说是以 $+\infty$ 为上界,它的上确界也是 $+\infty$. 这种语言在涉及集合、测度、积分时用得多一些.本书偶尔也会用一点.

# §2　度量空间和赋范线性空间

**1. 度量空间的基本性质**　上一节中我们详尽地讨论了实数理论,其目的不只在于给微积分理论中的一些基本定理以严格的证明.更重要的是,建立实数理论的途径为数学中许多其它问题提供了一个框架. 如果我们认为 $\mathbf{R}^n$ 的线性结构已经很清楚(这是线性代数的任务),在 $\mathbf{R}^n$ 上的极限理论也很清楚,则我们看到,微积分理论的这一个框架实际上包含了两种成分:一是它有线性空间结构,二是它有拓扑结构——在下一节明确定义什么是拓扑结构以前,我们就把它理解为有关极限的理论好了.而且这个拓扑结构是以两点之间的距离为基础的.但是,数学对象中具有这两种成分的可以说比比皆是.所以,对 $\mathbf{R}^n$ 作了详细讨论以后就可以仿照它来讨论一大类具有这两种结构的对象.这样一方面使我们进入了许许多多新的数学领域,另一方面又可以加深对微积分的理解.

首先,我们讨论以距离为基础的拓扑结构. 我们总是要讨论某一个集合 $X$,什么是距离呢?就是

**定义 1**　若对 $X$ 中任两个元 $x,y$ 均存在一个非负实数 $\rho(x,y)$ 具有以下性质,$X$ 就称为一个度量空间,而 $\rho(x,y)$ 就称为 $x,y$ 两点的距离:

(i) $\rho(x,y)\geqslant 0$, $\rho(x,y)=0$ 当且仅当 $x=y$ 时;

(ii) $\rho(x,y)=\rho(y,x)$;

(iii) (三角形不等式) $\rho(x,z)\leqslant\rho(x,y)+\rho(y,z)$.

$X$ 的子集 $X_1\subset X$ 中自然也赋有了以上的距离,且有同样的性质.所以 $X_1$ 也是一个度量空间,称为 $X$ 的子空间.以下我们都采取这样的用语:当我们说到一个空间时,时常不只指一个集合,而且是具有某种结构的集合.所谓子空间也不只是一个子集合,而是具有同一种结构(由原集合带来的结构)的子集合. 例如连续函数集合是可积函数集合的子集合,但当它们带有不同定义的距离时,就不能说前者是后者的子空间.

$\mathbf{R}^n$ 中的距离是由一点到原点的距离决定的.这就是初等几何中的平行四边形法则——对角线向量等于终点向量减去起点向量所决定的向量之长度.在这里 $\mathbf{R}^n$ 的线性结构起了作用:只有

在有了线性结构时才能谈到向量、向量之差、原点等等.对于一般具有线性结构的空间 $X$ 也是一样.

**定义 2**　设 $X$ 是一线性空间，而对其每一点 $x$ 都有一个非负实数 $\|x\|$ 适合以下条件，则 $\|x\|$ 称为 $x$ 的范数，而 $X$ 称为赋范线性空间.

(i) $\|x\|\geqslant 0$，$\|x\|=0$ 当且 $x=0$ 时；

(ii) $\|\lambda x\|=|\lambda|\,\|x\|$，$\lambda$ 为 $X$ 之系数域中之元；

(iii) (三角形不等式)：$\|x+y\|\leqslant\|x\|+\|y\|$.

**定理 1**　赋范线性空间 $X$ 在定义 $\rho(x,y)=\|x-y\|$ 后成一度量空间.

证明很容易，所以略去.

我们在前几章已多次见到过度量空间和赋范线性空间之例.最简单的当然是 $\mathbf{R}^n$.但即令如此简单的空间，也还有一点话可讲.$\mathbf{R}^n$ 中两点，$x=(x_1,\cdots,x_n)$ 和 $y=(y_1,\cdots,y_n)$ 的距离 $\|x-y\|$ 有多种方式定义.当然最常见的是欧几里得距离：

$$\|x-y\|_E=\Big[\sum_{i=1}^{n}(x_i-y_i)^2\Big]^{1/2}. \tag{1}$$

此外还有例如

$$\|x-y\|_1=\sum_{i=1}^{n}|x_i-y_i|,$$

$$\|x-y\|_2=\sup_i|x_i-y_i|$$

等等.但是不论怎么定义向量之范数，这些定义都是等价的.等价的意义解释如下：如果以某种方式定义了范数 $p(x)$，我们又规定 $\|\cdot\|_E$ 表示(1)所定义的范数，则

$$p(x)=\|x\|_E\,p(y),\quad y=x/\|x\|_E. \tag{2}$$

故当 $\|x\|_E\to 0$ 时 $p(x)\to 0$，由三角形不等式易见

$$|p(x)-p(x_0)|\leqslant p(x-x_0),$$

因此当 $\|x-x_0\|_E\to 0$ 时，$p(x)\to p(x_0)$，所以不论如何定义 $x$ 之范数，这个范数一定是 $x$ 在欧几里得度量意义下的连续函数.现在来看(2)，显然 $\|y\|_E=1$，所以 $y$ 位于 $\mathbf{R}^n$ 在欧几里得意义下的单位球面 $S^{n-1}$ 上(这里上标 $n-1$ 表示球面的维数，而不是 $x$ 的维数：

$$S^{n-1}:x_1^2+\cdots+x_n^2=1.$$

这个记号在整个数学中是通用的.)而 $p(y)$ 是 $S^{n-1}$ 上的连续函数.$S^{n-1}$ 是一个紧集，读者在上一节中已经熟知了 $\mathbf{R}$ 的有界闭区间是紧的，单位圆周在极坐标中就是闭区间 $[0,2\pi]$，自然也是紧的.$n-1$ 维单位球面在广义球坐标中是 $[0,2\pi]\times\cdots\times[0,2\pi]\times[0,\pi]$，也容易看到是紧的，而连续函数在紧集上一定会达到最大值 $M$ 和最小值 $m$：

$$m\leqslant p(y)\leqslant M,\qquad y\in S^{n-1}.$$

$m$ 一定不会是 0，因为 $p(y)$ 必在某一点达到它：$p(y_0)=0$，而由范数定义又有 $y_0=0$.但是 $S^{n-1}$ 上是不会有 0 的.代入(2)即知，不论 $p(x)$ 是什么样的范数，必有

$$m\|x\|_E\leqslant p(x)\leqslant M\|x\|_E,\qquad m>0. \tag{3}$$

当然由此也就有

$$\lambda p(x)\leqslant\|x\|_E\leqslant\Lambda p(x),\ 0<\lambda<\Lambda.$$

两个范数如果适合这样的不等式就说是等价的.所以,$\mathbf{R}^n$ 上一切范数都是等价的.但是,我们下面只讨论欧几里得范数 $\|x\|_E$,因为只有利用它,才能最方便地把种种几何性质赋给 $\mathbf{R}^n$.所以以后凡说到 $\mathbf{R}^n$ 中的范数,若无特别申明,总是指的这个范数,而 $\|\cdot\|_E$ 的下标 $E$ 以及诸如"欧几里得意义下的"等限制语,如无特殊需要都一律不用.

既然讲了 $\mathbf{R}^n$,读者会问到 $\mathbf{C}^n$. 应该指出,$\mathbf{C}$ 可以分成实、虚部,而由两个 $\mathbf{R}$ 合成. 这当然只是一个形式的处理方法,而没有涉及更深层的问题.但因本书范围不许可我们进入这个领域,所以本书一直是采取这个形式的处理法,以下也是一样.

次于 $\mathbf{R}^n$,我们还见到了许多函数空间. 首先是某有界闭区间 $[a,b]$上的连续函数空间. 现在应该考虑紧空间 $K$ 上的连续函数之空间 $C(K)$,而不限于 $K$ 为 $[a,b]\subset\mathbf{R}$. 证明是完全一样的,$C(K)$是线性空间是自明的,它也是赋范的,因为对任一 $f\in C(K)$,可以证明

$$\|f\|=\sup_K|f(x)|$$

就是一个范数.首先,它是有意义的,因为定义在一个紧集上的连续函数一定有界而且一定可以达到最大最小值(所以(3)式也可以写为 $\|f\|=\max\limits_K|f(x)|$),定理 8 之(i)当 $K=[a,b]$时给出了证明.但实际上这个证明对任意紧集 $K$ 适用,因为它只用了$[a,b]$可以用有限多个小区间覆盖这一事实. 对于一般的 $K$,只要把小区间改成适当的邻域,再用紧性定义即可. 下面好几个性质我们也会不加证明地用于 $K$,理由同此. 不过下面我们还是会详细证明的. 范数的三个性质中(i),(ii)是自明的,(iii)则很容易,因为

$$|f(x)+g(x)|\leqslant|f(x)|+|g(x)|,\forall x\in K,$$

先在右方取最大值,对于一切 $x\in K$

$$\begin{aligned}|f(x)+g(x)|&\leqslant\sup_K(|f(x)|+|g(x)|)\\&\leqslant\sup_K|f(x)|+\sup_K|g(x)|=\|f\|+\|g\|.\end{aligned}$$

左方再取一次最大值,即有

$$\|f+g\|\leqslant\|f\|+\|g\|.$$

但是,我们不能先在左方取最大值,也不能左、右方同时取最大值,因为这样做,$\leqslant$关系无法保持.

其次要看可积函数空间 $L^p(I)$,$1\leqslant p<+\infty$,这时我们可以定义 $f\in L^p$ 之范数为

$$\|f\|_{L^p}=\left(\int_I|f(x)|^p\mathrm{d}x\right)^{1/p}. \tag{4}$$

这里 $I$ 既可以是有限区间$[a,b]$,也可以是无限区间以至于整个 $\mathbf{R}$,这都与 $I$ 是否为紧无关. 所以下面记为 $L^p$.

范数的三个要求(i)至(iii)中,(iii)就成了闵可夫斯基不等式

$$\|f+g\|_{L^p}\leqslant\|f\|_{L^p}+\|g\|_{L^p}.$$

值得注意的是(i):若 $\|f\|_{L^p}=0$,我们只能得到 $f(x)$几乎处处为 0. 在第四章中我们强调指出过,在勒贝格积分理论中,几乎处处为 0 的函数就可以忽略不计.现在我们再强调一次说,一个勒贝格可积函数其实是指一个等价类:若 $f,g\in L^p$,而且 $f(x)-g(x)=0$ 几乎处处成立,就说 $f\sim g$是互相等价的.每一个等价类就算一个可积函数.特别是 $f(x)=0$(处处为 0)所在的等价类,即一切几乎处处为 0 的勒贝格可积函数之类,亦即零类,即是 $L^p$ 中的零元.

要注意 $L^\infty$,其中的范数是

$$\|f\|_{L^\infty}=\text{ess sup}|f(x)|.$$

这与 $C(K)$中的范数是“一样”的，所以对于紧集 $K$，$C^\infty(K)\subset L^\infty(K)$. $C^\infty(K)$有许多重要性质在 $L^\infty$ 中也都成立，但是终究它们是不相同的空间.

$L^p$ 空间中 $L^2$ 最为重要. 它属于这样一种赋范空间，即其范数是由内积产生的，即是说，对于这个空间的任意两个元 $f,g$，一定有一个实(复)数 $\langle f,g\rangle$，$((f,g))$与之对应，$\langle f,g\rangle$ (或$(f,g)$)分别有以下的性质：

**定义 3**　设 $H$ 为一实的(或是复的)线性空间，若对任意 $f,g\in H$ 可以定义一个实数(或复数) $\langle f,g\rangle$ (或$(f,g)$)而具有以下性质：

(i) $\langle f_1+f_2,g\rangle=\langle f_1,g\rangle+\langle f_2,g\rangle$，或$(f_1+f_2,g)=(f_1,g)+(f_2,g)$；

$\langle f,g_1+g_2\rangle=\langle f,g_1\rangle+\langle f,g_2\rangle$或$(f,g_1+g_2)=(f,g_1)+(f,g_2)$；

(ii) $\langle\lambda f,g\rangle=\lambda\langle f,g\rangle$，$\lambda\in\mathbf{R}$，或$(\lambda f,g)=\lambda(f,g)$，$\lambda\in\mathbf{C}$；

$\langle f,\lambda g\rangle=\lambda\langle f,g\rangle$，$\lambda\in\mathbf{R}$，或$(f,\lambda g)=\overline{\lambda}(f,g)$，$\lambda\in\mathbf{C}$；

(iii) $\langle f,g\rangle=\langle g,f\rangle$，或$(f,g)=\overline{(g,f)}$；

(iv) $\langle f,f\rangle\geqslant0$，$\langle f,f\rangle=0$，当且仅当 $f=0$；或$(f,f)\geqslant0$，$(f,f)=0$ 当且仅当 $f=0$.

则 $\langle f,g\rangle$ (或$(f,g)$)称为内积，$H$ 称为实(或复)内积空间.

实的内积又称为欧几里得内积或欧几里得配对；复的内积又称为埃尔米特内积或埃尔米特配对. 具有这些性质的内积(即配对)，按以下方式在这个空间中定义范数

$$\|f\|=[\langle f,f\rangle]^{\frac{1}{2}}\text{或}\|f\|=[(f,f)]^{\frac{1}{2}}. \tag{5}$$

由性质(iv)知道这样定义的 $\|f\|$ 是有意义的，它是一个非负实数，而且适合范数所应具有的性质，即定义 1 中的(i)～(iii). (i)～(ii)容易证明，关键是要证明三角形不等式(iii)，为此先要证明关于内积的一个重要不等式.

对于实内积空间，令 $\lambda$ 为任意实数，我们来考虑内积 $\langle f+\lambda g,f+\lambda g\rangle$，由(iv)它是非负的. 另一方面

$$\begin{aligned}0\leqslant\|f+\lambda g\|^2&=\langle f,f\rangle+2\lambda\langle f,g\rangle+\lambda^2\langle g,g\rangle\\&=\|f\|^2+2\lambda\langle f,g\rangle+\lambda^2\|g\|^2.\end{aligned}$$

式右是 $\lambda$ 的二次三项式，它恒非负这一事实等价于其判别式恒非正，即有

$$|\langle f,g\rangle|\leqslant\|f\|\ \|g\|. \tag{6}$$

(6)式称为施瓦茨不等式.

对复内积空间，证明比较曲折. 首先设$(f,g)\neq0$，它可写为$(f,g)=|(f,g)|e^{i\theta}$而 $\theta$ 有定义，令 $\lambda=te^{i\theta}$，$t$ 为实数，并考虑内积$(f+\lambda g,f+\lambda g)$，它也是恒非负的. 化简后有

$$\begin{aligned}0\leqslant\|f+\lambda g\|^2&=(f,f)+\overline{\lambda}(f,g)+\lambda\overline{(f,g)}+|\lambda|^2(g,g)\\&=\|f\|^2+2t|(f,g)|+t^2\|g\|^2.\end{aligned}$$

这是一个 $t$ 的恒非负的二次三项式且 $t$ 为实数，于是又有

$$|(f,g)|\leqslant\|f\|\cdot\|g\|. \tag{$6_1$}$$

其次看$(f,g)=0$ 的情况，这时上式自然成立. ($6_1$)式也称为施瓦茨不等式.

有了这个重要的

$$\begin{aligned}\| f+g \|^2 &= (f+g,f+g) = (f,f)+2\mathrm{Re}(f,g)+(g,g)\\ &\leqslant \| f \|^2+2|(f,g)|+\| g \|^2\\ &\leqslant (\| f \|+\| g \|)^2.\end{aligned}$$

内积空间是最重要的赋范线性空间. 它是 $\mathbf{R}^n$ 的直接推广. 实际上 $\mathbf{R}^n$ 也是内积空间(但这时一定要对 $\mathbf{R}^n$ 赋以欧几里得范数 $\| x^2 \| = \sum_{i=1}^{n} |x_i|^2$ 才行), 因为对于 $x, y \in \mathbf{R}^n$ 我们可定义内积为

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i .$$

$L^2(I)$, $(I\subset\mathbf{R}^n)$也是内积空间, 其中的内积当我们限于考虑实值函数时为

$$\langle f,g\rangle = \int_I f(x)g(x)\mathrm{d}x.$$

在考虑复值函数时则为

$$(f,g) = \int_I f(x)\overline{g(x)}\mathrm{d}x.$$

这个空间有许多重要性质在第四章中介绍过. 我们现在要提醒, 其完备性应特别注意. 如果在内积空间中再引入完备性, 将得到所谓希尔伯特空间. 它是数学物理中应用最为广泛的空间, 而 $L^2(I)$就是它的典型(这句话的意思也将在以后明确).

以 $L^2$ 空间为基础的索伯列夫空间 $H^k(I) = \{f(x); \mathrm{D}^\alpha f(x)\in L^2$, 当 $|\alpha|\leqslant k$ 时$\}$也是非常重要的内积空间. 这里的 $\mathrm{D}^\alpha f(x)$是指 $f(x)$之阶数为$|\alpha|$的广义函数意义下的导数. 这个空间中的内积及范数可以定义如下:

$$(f,g) = \int_I \sum_{|\alpha|=0}^{k} \mathrm{D}^\alpha f(x)\cdot\overline{\mathrm{D}^\alpha g(x)}\mathrm{d}x,$$

$$\| f \|^2 = \int_I \sum_{|\alpha|=0}^{k} |\mathrm{D}^\alpha f(x)|^2\mathrm{d}x.$$

第四章中证明了, $H^k(I)$在这个范数下也是完备的. 它也是希尔伯特空间.

本书中除了种种函数空间外还介绍了一些数列空间. 例如在傅里叶级数一节中就指出, 当 $f(x)\in L^1([0,2\pi])$时, 我们可以用它的傅里叶系数$\{C_n\}$来刻画 $f(x)$. 这时

$$C_n = \frac{1}{2\pi}\int_0^{2\pi} f(x)\mathrm{e}^{-\mathrm{i}nx}\mathrm{d}x,\ n=0,\pm 1,\pm 2,\cdots.$$

那里我们指出, 这个数列空间很难于刻画, 而这正表现了傅里叶级数理论的重大困难. 但不论如何, 我们知道了, 数列是描述物理现象的重要手段, 因此它们也是重要的研究对象. 最常见到的数列空间是 $l^p$, $1\leqslant p\leqslant+\infty$, 它们是与 $L^p$ 空间相应的. 当 $p<+\infty$时

$$l^p = \{c=(c_1,c_2,\cdots,c_2,\cdots),\ \sum_{n=1}^{\infty} |c_n|^p<+\infty\},$$

其中 $c_1,\cdots,c_n,\cdots$是复数. $c$ 的范数定义为

$$\| c \|_{l^p} = \left(\sum_{n=1}^{\infty} |c_n|^p\right)^{1/p}.$$

它所应满足的三角形不等式就是著名的闵可夫斯基不等式: 对于 $c=(c_1,\cdots,c_n,\cdots)\in l^p$ 和 $d=$

$(d_1,\cdots,d_n,\cdots)\in l^p$ 有

$$\|c+d\|_{l^p}\leqslant\|c\|_{l^p}+\|d\|_{l^p}.$$

当 $p=+\infty$时，$l^p=l^\infty$就规定是有界数列空间：

$$l^\infty=\{c=(c_1,\cdots,c_n,\cdots),\sup_n|c_n|<+\infty\}.$$

在其中我们也可以定义范数为

$$\|c\|_{l^\infty}=\sup_n|c_n|.$$

$l^2$ 是特别重要的，因为它是内积空间. 对 $c,d\in l^2$，我们定义其内积如下：当 $c,d$ 是实数列时

$$\langle c,d\rangle=\sum_{n=1}^{\infty}c_nd_n.$$

当 $c,d$ 为复数列时，

$$(c,d)=\sum_{n=1}^{\infty}c_n\overline{d}_n.$$

这些级数的收敛性都可以由 $c,d\in l^2$，以及

$$\left|\sum_{j=1}^{k}c_{n+j}\overline{d}_{n+j}\right|\leqslant\sum_{j=1}^{k}|c_{n+j}\overline{d}_{n+j}|$$
$$\leqslant\Big(\sum_{j=1}^{k}|c_{n+j}|^2\Big)^{1/2}\Big(\sum_{j=1}^{k}|d_{n+j}|^2\Big)^{1/2}$$

得到. 上面的不等式就是关于数列的施瓦茨不等式应用于级数 $\sum c_n\overline{d}_n$ 的一段.

$l^2$ 和 $L^2$ 一样，都具有完备性，因此都是希尔伯特空间.

**2. 度量空间的拓扑性质**　以上我们列举了一些在本书中介绍到而且在数学中很有用的赋范线性空间. 但是我们完全没有涉及有关极限，连续性等拓扑问题，而只讨论了这些空间的线性结构. 现在要进一步讨论拓扑问题时，我们又必须以 **R** 中建立拓扑结构的方法为模式. 在 **R** 中研究连续性时，两点的距离起了重要作用，例如讨论由 $(a,b)\subset\mathbf{R}$ 到 **R** 中的连续映射 $f(x)$，如果 $f$ 映 $x_0\in(a,b)$ 到 $f(x_0)\in\mathbf{R}$. 这一段话，在现代数学中时常以很简单的公式表示为：

$$f:(a,b)\subset\mathbf{R}\to\mathbf{R},x_0\mapsto f(x_0).$$

$f(x)$在 $x_0$ 点连续的定义是：对任一 $\varepsilon>0$ 必有 $\delta(\varepsilon)>0$ 存在，而当

$$|x-x_0|<\delta\text{ 时}|f(x)-f(x)|<\varepsilon.$$

这里用的是 $x$ 与 $x_0$ 在 $x$ 轴(它是 **R**，常记为 $\mathbf{R}_x$)上的距离，以及 $f(x)$与 $f(x_0)$在 $y$ 轴(另一个 **R**，常记为 $\mathbf{R}_y$)上的距离. 把这些概念推广到一般的度量空间上去是很容易的事. 但为此，我们要介绍一些名词或概念：我们把度量空间 $X$ 中的以 $x_0$ 为心，$r$ 为半径的开球$\{x\in X;\rho(x,x_0)<\varepsilon\}$，记为 $\mathring{B}_\varepsilon(x_0)$；闭球$\{x\in X;\rho(x,x_0)\leqslant\varepsilon\}$记为 $B_\varepsilon(x_0)$. 这里开与闭的说法，以及用上加一个小圆圈表示“开”，都将与以后的定义一致，但在目前还未讨论开、闭之一般概念前，它们就只是一些名词与记号，我们暂时不必讨论它们的意义.

**定义 4**　若 $X$ 为一度量空间(其度量为 $\rho(x,y)$)，$U$ 为 $X$ 之子集，若对一切 $x_0\in U$，均有

$x_0$

按照这样的定义,开球确实是开集.事实上,若 $x_1\in \overset{\circ}{B}_\varepsilon(x_0)$,则 $\rho(x_1,x_0)<\varepsilon$.现在取 $\varepsilon_1<\varepsilon-\rho(x_1,x_0)$,则对 $\overset{\circ}{B}_{\varepsilon_1}(x_1)$中任一点 $x^*$,由三角形不等式.

$$\rho(x^*,x_0)\leqslant\rho(x^*,x_1)+\rho(x_0,x_1)<\varepsilon_1+\rho(x_0,x_1)<\varepsilon.$$

从而 $\overset{\circ}{B}_{\varepsilon_1}(x_1)\subset\overset{\circ}{B}_\varepsilon(x_0)$,所以 $\overset{\circ}{B}_\varepsilon(x_0)$是开集.总之,“开球”之“开”和“开集”之“开”是一回事.

用同样的方法可以证明$\{x\in X,\rho(x,x_0)>\varepsilon\}$也是开集,所以它的余集 $B_\varepsilon(x_0)$是闭集.即是说,闭球确实是一个闭集. 定义中指 $\overset{\circ}{B}_\varepsilon(x_0)$是 $x_0$ 的邻域(其一般定义见下文),其实所谓 $x_0$ 的邻域定义是包含了一个以 $x_0$ 为心的开球的集合,至于它本身则可开也可闭.所以时常也说到开邻域与闭邻域.但有一点很清楚,任意含有 $x_0$ 的开集都是 $x_0$ 的开邻域.

关于开集与闭集有以下的基本定理.

**定理 2** (i) $X$ 与$\varnothing$均为开集.

(ii) 任意多个开集$\{U_\lambda\}$之并$\bigcup\limits_\lambda U_\lambda$ 仍为开集.

(iii) 有限多个开集之交仍是开集.

**证** (i) $X$ 为开是明显的.至于空集也是开集是可以证明的,但因涉及到一些逻辑方面的问题解释起来要费一些功夫,我们把它当作一个约定也可以.

(ii) 若 $x_0\in\bigcup\limits_\lambda U_\lambda$,必有某个 $\lambda_0$,使 $x_0\in U_{\lambda_0}$,所以由开集的定义知,$x_0$ 有一个邻域 $U\subset U_{\lambda_0}\subset\bigcup\limits_\lambda U_\lambda$,所以$\bigcup\limits_\lambda U_\lambda$ 是开集.

(iii) 若 $x_0\in\bigcap\limits_{k=1}^N U_k$,则 $x_0$ 属于每个 $U_k$,从而有开球 $\overset{\circ}{B}_{\varepsilon_k}(x_0)\subset U_k$.在有限多个正数 $\varepsilon_k$ 中,必有一个最小的 $\varepsilon_0$ 仍为正数,而对一切 $k$,$\overset{\circ}{B}_{\varepsilon_0}(x_0)\subset\overset{\circ}{B}_{\varepsilon_k}(x_0)\subset U_k$,所以 $\overset{\circ}{B}_{\varepsilon_0}(x_0)\subset\bigcap\limits_{k=1}^N U_k$,而知$\bigcap\limits_{k=1}^N U_k$ 也是开集.

利用开集之余集即为闭集,就可以得到关于闭集的相应基本定理.

**定理 3** 度量空间 $X$ 的闭集有以下的性质:

(i) $X$ 与$\varnothing$均为闭集.

(ii) 任意多个闭集$\{V_\lambda\}$之交$\bigcap\limits_\lambda V_\lambda$ 仍为闭集.

(iii) 有限多个闭集之并仍为闭集.

**证** 考虑闭集之余集,本定理即归结为定理 2.

**注 1** 最容易引起误会是全空间 $X$ 与空集$\varnothing$既是开集又是闭集,岂非矛盾?我们要问为什么不能既是此又是彼?所谓矛盾,由何说起?实际上,说一个对象既是什么又不是什么,这才是矛盾,“$A$ 是开集”与“

开和闭区间分别是开集与闭集.现在取可数多个开集 $I_h=\left(a-\frac{1}{h},b+\frac{1}{h}\right)$，$a<b$，很明显它们的交 $\bigcap_{h=1}^{\infty} I_h=[a,b]$是闭集而非开集. 又看可数多个闭集 $J_h=\left[a+\frac{1}{h},b-\frac{1}{h}\right]$，$b-a>0$，则 $\bigcup_{h=h_0}^{\infty}=(a,b)$是开的而非闭的. $h$ 要适合$\frac{2}{h}<(b-a)$，以保证 $a+\frac{1}{h}$在$b-\frac{1}{h}$之左.

有了开集和闭集的概念后，就可以进一步介绍其它一些重要的概念，首先是邻域的概念.

**定义 5** 空间 $X$ 中一点 $x$ 的邻域 $U$，即 $X$ 的这样一个子集：不但 $x\in U$，还有一个含 $x$ 的开集 $V$ 也含于 $U$ 内：$V\subset U$.

我们这里讲空间 $X$ 是指度量空间.但是下一节我们还会介绍另一类更广泛的空间：拓扑空间，其中一点之邻域的定义仍然与上逐字相同，所以我们就只讲“空间”而不特指度量空间. 下面的定理 4~5，定义 6~7 也有类似情况，请读者注意.

对于一点的邻域 $U$，我们只要求它包含一个开集 $V$ 而 $x$ 属于 $V$，而不要求本身为开，这样做在讨论一些数学问题上会带来方便.如果 $U$ 本身也是开集或闭集，就称它为开或闭邻域.开邻域特别值得注意.因为开集 $U$ 是其中任何一点的开邻域——我们只需把定义 5 中的 $V$ 取为 $U$ 本身也可以看出来了.许多书上讲邻域总是指的开邻域，而且直接就说，一点 $x$ 的邻域就是含 $x$ 的任一开集，这样做在许多问题上会带来方便.

我们把一点 $x$ 的所有邻域之集合记作 $N_x$，称为 $x$ 点之邻域系. $N_x$ 是 $X$ 的子集族，是“集合的集合”，它的元素不是 $X$ 之点，而是 $X$ 的一种子集 $U$：$U$ 是 $x$ 之邻域. $N_x$ 自然不是空的，就是说，每一点 $x$ 都有邻域，这当然是平凡不足道的，因为全空间 $X$ 是任一点的邻域，这时只要取 $U=V=X$ 即可用定义 5 来验证 $X$ 是 $x$ 的邻域，关于邻域的基本性质，我们有下面的

**定理 4** 空间 $X$ 中任一点的邻域系 $N_x$ 有以下性质：

(i) 若 $U_\lambda\in N_x$，则 $\bigcup_{\lambda} U_\lambda\in N_x$；

(ii) 若 $U_1,\cdots,U_N\in N_x$，则 $\bigcap_{k=1}^{N} U_k\in N_x$；

(iii) 若 $U\in N_x$，

(ii) $x$ 称为 $A$ 之外点，如果 $x$ 是 $A$ 之余集 $A^c$ 之内点；

(iii) 若 $x$ 既非 $A$ 之内点，又非 $A$ 之外点，就说 $x$ 是 $A$ 之边界点.

很明显，若 $x$ 是 $A$ 之内点则 $x\in A$（因为 $x$ 属于它自己的任一邻域）；若 $x$ 是 $A$ 之外点，则 $x\notin A$. 若 $x$ 是 $A$ 之边界点，则既可能 $x\in A$，也可能 $x\notin A$.

我们还可以这样来描述内点、外点与边界点：如果 $x$ 有某一开邻域全含于 $A$ 中，则 $x$ 是内点，这一点已在(i)中表述了. 若 $x$ 有某一开邻域完全不在 $A$ 内，即与 $A$ 完全不相交，则 $x$ 是外点. 若 $x$ 之任一开邻域中都既有 $A$ 中之点，也有 $A$ 外之点，则 $x$ 是边界点. 请读者注意现在"某一"改成了"任一"，如果仍然沿用"某一"就一定会错. 例如设 $A$ 是 $(x, y)$ 平面上的单位圆：$x^2+y^2\leqslant 1$，由定义适合 $x^2+y^2 \overset{<}{\underset{>}{=}} 1$ 的 $(x, y)$ 分别是内点、边界点和外点. 但是内点 $(0,0)$ 有"某一"邻域 $U$（例如以 $(0,0)$ 为心，2 为半径的圆），其中既有 $A$ 中之点又有 $A$ 外之点. 还要注意，边界点的"边界"二字与我们直观的理解很不相同. 例如，设在 $A$ 中除去一个内点 $x_0$，并记 $A_1=A\setminus\{x_0\}$，则 $x_0$ 成了 $A_1$ 的边界点，因为它的任一邻域中，既有 $A_1$ 中之点（这是明显的）也有 $A_1$ 以外的点，即 $x_0$ 本身. 同样，若取 $A$ 的一个外点 $x_0$ 附加到 $A$ 上成为 $A_1=A\cup\{x_0\}$，$x_0$ 现在是 $A_1$ 的一个孤立点了，但是它也是 $A_1$ 之边界点，因为在它的任一邻域中都既有 $A_1$ 中之点，即例如 $x_0$ 本身，也有 $A_1$ 外的点，即与 $x_0$ 充分接近而又与 $x_0$ 不同的点. 这里用"充分接近"的说法，一方面可以从度量的意义来理解它，即有一个很小的 $\delta>0$，而

$$0<\rho(x, x_0)<\delta.$$

但我们上面又说了，这里的讲法应该对于一般的拓扑空间也适用，所以详细地讲应该是：既然 $x_0$ 是 $A$ 之外点，则必有 $x_0$ 的一个邻域 $U\cap A=\varnothing$，$U$ 中与 $x_0$ 不同的点就是上面说的"与 $x_0$ 充分接近而又与 $x_0$ 不同的点".

边界点中既有这些麻烦事，则它与例如格林定理，柯西定理中讲的某区域的边界就不相同了. 那里讲的某区域的边界，例如一条曲线、一个曲面，我们应该换一个名词，称为边缘. $A$ 之边界点之集合记为 $\mathrm{bd}A$，$A$ 的边缘记为 $\partial A$. 但是它们的英文术语是相同的即为 boundary，这里作这样的划分并非过分谨慎，而是有重要意义的. 关于边缘的问题我们将在下一章详细讨论.

从内点，外点和边界点概念可以给出下面几个重要概念

**定义 7 和定理 5** 设 $A$ 为空间 $X$ 之子集，则我们称 $A$ 之内点的集合为 $A$ 之内域，记为 $\mathring{A}$ 或

$\bigcup_{\lambda} U_{\lambda} \subset \mathring{A}$，比较这两个结果即得 $\mathring{A} = \bigcup_{\lambda} U_{\lambda}$.

取 $A$ 之余集 $A^c$ 并作以上推理即得关于 $\overline{A}$ 的相应结论.关于 $\mathrm{bd}A$ 的结论自明. 证毕.

由这个定理即得 $\mathring{A}$ 是开集，$\overline{A}$ 是闭集，而 $\mathrm{bd}A = X \setminus (\mathring{A} \cup \mathrm{int}(A^c))$ 是闭集，而且还看到

$$\overline{A} = \mathring{A} \cup \mathrm{bd}A = A \cup \mathrm{bd}A,$$

$$\mathring{A} = \overline{A} \setminus \mathrm{bd}A = A \setminus \mathrm{bd}A.$$

这些结论的证明都略去了.

再提一下，定义 7 和定理 5 也适用于一般的拓扑空间.下面我们讨论有关极限和连续性的问题却只限于度量空间.在通常的微积分教程中讨论极限和连续问题时，基本上是讨论数列与函数的问题.说是“基本上”是因为有重要的例外，如积分和的“极限”就不是这两种类型.对于一般的度量空间 $X$，我们暂时也只限于两种情况，即序列 $\{x_n\}$，$x_n \in X$（现在我们不说数列，因为 $x_n$ 一般地并不是数）和由一个度量空间 $X$ 到另一个度量空间 $Y$ 的映射.我们先讨论后一个情况.设 $X$ 和 $Y$ 中的度量分别是 $\rho_X$ 和 $\rho_Y$.若 $f: X \to Y$，$x_0 \mapsto y_0$，我们说 $f$ 在 $x_0$ 连续即是对任一 $\varepsilon > 0$，必可找到 $\delta(\varepsilon) > 0$ 使当 $\rho_X(x, x_0) < \delta$ 时，$\rho_Y(f(x), y_0) < \varepsilon$.这里要注意，若 $f(x) = y$，我们说 $y$ 是 $x$（在 $f$ 下）的像，而 $x$ 是 $y$ 的原像（pre-image）.$X$ 中的每一个 $x$ 在 $f$ 下只有一个像点 $f(x)$ 在 $Y$ 中，但 $Y$ 中一点 $y$ 在 $f$ 下的原像则不一定只有一个点.这与函数 $y = f(x)$ 必须是单值的，但同一个 $y$ 却可以有许多不同的 $x$ 与之相应的是一样的.如果同一个 $y$ 原像只有一点，则我们会得到由 $Y$ 中 $f$ 的值域（下面暂记为 $Y$，其实不准确）到 $X$ 的另一个映射 $f^{-1}$，$Y \to X$，$f^{-1}$ 称为逆映射，亦即反函数.但在 $y$ 之原像多于一个点时，我们时常也使用 $f^{-1}(y)$ 这个记号.这时，它表示一个集，其中的点，而且在此集外再无其它点，被 $f$ 映到 $y$.回到连续性问题，上述“$\varepsilon$—$\delta$ 定义”就是说：$Y$ 中的球 $\rho_Y(y, y_0) < \varepsilon$ 的原像必在 $X$ 的某个球 $\rho_X(x, x_0) < \delta$ 内. 这当然不是说，球 $B_\varepsilon(y_0)$ 中所有的点都是球 $B_\delta(x_0)$ 中某点的像.而是说，若球 $B_\varepsilon(y_0)$ 中一点，如果在映射 $f$ 的值域中的话（当然，这就意味着 $B_\varepsilon(y_0)$ 中有些点可能根本不在 $f$ 的值域中.如果对这种点也要使用 $f^{-1}(y)$ 的记号的话，则 $f^{-1}(y) = \varnothing$），则这个点一定有原像在 $B_\delta(x_0)$ 中（当然，这不是说所有的原像点都只在 $B_\delta(x_0)$ 内而不能在其外，这里只说在 $B_\delta(x_0)$ 内有原像之点，而且 $B_\delta(x_0)$ 必被 $f$ 映入 $\rho_Y(y, y_0) < \varepsilon$ 内）.我们看看函数 $y = f(x)$ 的两个例子就明白了.常值函数 $y = y_0$，即 $f(x) \equiv y_0$，当然在任一点 $x_0$ 都连续.但是 $\rho_X(x, x_0) < \delta$ 即 $|x - x_0| < \delta$ 之像却只是 $\rho_Y(y, y_0) < \varepsilon$ 即 $|y - y_0| < \varepsilon$ 中一个点 $y_0$，而不是整个区间.又如 $y = x^2$ 在 $x = 2$ 处连续，而 $x = 2$ 的像是 $y = 4$，但是 4 附近的 $y$ 点之原像除了有一个在 2 附近以外，还有一个在 $-2$ 附近.从上面讨论的例子中可以看到，度量在其中并不起作用，我们可以这样来表述映射 $f: X \to Y$，$x_0 \mapsto y_0$ 点处的连续性如下：任取 $y_0$ 在 $Y$ 中的一个邻域 $V$，必有 $x_0$ 在 $X$ 中一个相应的邻域 $U$，被 $f$ 映入 $V$ 内：$f(U) \subset V$.所以，一不是说 $f(U)$ 可以把 $V$ 填满，或说 $f$ 把 $U$ 映到 $V$ 上，二不是说，$U$ 以外没有其它的点也被 $f$ 映入 $V$ 内. 同样可以考虑 $f: X \to Y$ 在 $X$ 上连续的定义.当然会问如果只需讨论 $f$ 在 $X$ 的一个子集合上连续该怎么办？这一点将在下一节回答.现在我们说 $f: X \to Y$ 在 $X$ 上连续的充分必要条件是：$Y$ 中每个开集 $V$ 在 $f$ 下的原像 $f^{-1}(V)$ 都是 $X$ 中的开集.这一点我们就不来

证明了.

读者很容易看到,关于映射连续性的讨论很容易推广到一般的拓扑空间中去,但是下面关于序列收敛的讨论却本质地只能限于度量空间.

**定义 8**　设 $X$ 为一度量空间,$\{x_n\}\subset X$.若有一点 $x_0\in X$,使得对任一 $\varepsilon>0$ 均有 $N(\varepsilon)>0$ 存在,而当 $n>N$ 时,$\rho(x_n,x_0)<\varepsilon$,就说$\{x_n\}$收敛于 $x_0$,记作$\lim\limits_{n\to\infty}x_n=x_0$.

于是我们面临的问题当然是,$\{x_n\}$收敛于 $x_0$ 的条件是什么.当然我们就想到,应该把有关 $\mathbf{R}$ 的概念和结论都搬过来.于是有

**定义 9**　设 $X$ 为一度量空间,$\{x_n\}\subset X$,若对任一 $\varepsilon>0$ 均有 $N(\varepsilon)>0$ 存在,使当 $m,n>N$ 时,$\rho(x_m,x_n)<\varepsilon$,则称$\{x_n\}$为一柯西序列.

当然一般说来,$X$ 中的柯西序列不一定都是收敛的,即不一定以 $X$ 中某一点 $x_0$ 为极限.于是我们要区别出一类特别重要的度量空间,这就是完备的度量空间.

**定义 10**　若度量空间 $X$ 中的任一柯西序列都是收敛的,即以 $X$ 中某一点 $x_0$ 为极限,我们就说 $X$ 是完备的度量空间.

在本书中我们已看到一个度量空间是否完备是多么重要.在上一节中我们指出,有理数系 $\mathbf{Q}$(读者可以自行验证,$\mathbf{Q}$ 也是度量空间)是不完备的,而这导致微积分的许多定理得不到证明.于是我们把它完备化得到实数系 $\mathbf{R}$,这才把整个微积分学放在可靠的基础上.我们又在积分学一章中看到,黎曼可积函数空间(它也不只是度量空间,而且是赋范空间,以黎曼积分 $\int_a^b|f(x)|\mathrm{d}x$ 为范数 $\|f\|$.读者也不妨自行验证一下)也是不完备的,导致对于黎曼积分一般地不能在积分号下求极限.这正是黎曼积分真正的缺点.我们引入勒贝格积分,有了里斯-费希尔定理,知道 $L^p$ 是完备空间,这才克服了困难.这是整个分析数学极重要的一大步.因此我们要问,能否仿效构造 $\mathbf{R}$ 的方法,对一个并不完备的度量空间 $X_0$,构造出一个完备的度量空间 $X$ 来?这里我们要求 $X_0\subset X$,而且 $X_0$ 的两点 $x$ 和 $y$ 在 $X_0$ 中的距离 $\rho_{X_0}(x,y)$与它们在 $X$ 中的距离 $\rho_X(x,y)$相等,同时要求 $X_0$ 在 $X$ 中稠密,即任一 $x_0\in X$ 必为$\{x_n\}\subset X_0$ 在 $\rho_X$ 下的极限:$\rho_X(x_n,x_0)\to0$,亦即$\lim\limits_{n\to\infty}x_n=x_0$.对于度量空间这个问题有肯定的答案.

我们仍然仿照构造 $\mathbf{R}$ 的方法,取 $X_0$ 中一个柯西序列$\{x_n\}$,并把它当作一个新空间的元素.准确些说,如果 $X_0$ 有两个柯西序列$\{x_n\}$与$\{x_n'\}$,如果 $\rho_{X_0}(x_n,x_n')\to0$.就说这两个柯西序列等价:$\{x_n\}\sim\{x_n'\}$.我们把柯西序列的等价类当作一个元素.并记这些等价类的集合为 $X$.我们要证明 $X$ 是一个度量空间,而 $X_0$ 是它的子空间.但 $X_0$ 中之元素是一些点,$X$ 之元素是一些这样的点组成的柯西序列之等价类,怎么可能 $X_0$ 是 $X$ 的子空间呢?这里我们也仿照 $\mathbf{R}$ 中的作法.注意到序列$\{x_0,x_0,\cdots,x_0,\cdots\}$也是柯西序列,因为其中任两点之距离为 0:$\rho_{X_0}(x_0,x_0)=0<\varepsilon$.这个序列称为常驻序列.于是我们让 $x_0$ 表示该常驻序列的等价类.在这样的意义下 $X_0\subset X$.当然我们还要考虑 $X_0$ 中的距离与 $X$ 中的距离(如果确有距离的话)的关系,我们最终是要证明

**定理 6**(**完备化定理**)　设 $X_0$ 为一度量空间,则必可找到一完备的度量空间 $X$,使 $X_0$ 为其等距的子空间,即不但有 $X_0\subset X$,而且 $X_0$ 中任两点 $x,y$ 在 $X_0$ 中的距离与其在 $X$ 中的距离相

等：$\rho_{X_0}(x,y)=\rho_X(x,y)$，而且 $X_0$ 在 $X$ 中稠密. 这样的 $X$ 在等距意义下是唯一的，称为 $X_0$ 之完备化.

**证**　上面我们已作出了 $X_0$ 中之柯西序列的等价类之集合 $X$，现在在 $X$ 中赋予距离如下：

设有 $X$ 中之两点，即 $X_0$ 中两个柯西序列之等价类 $x$ 与 $y$，在这两个等价类中各取一个代表元 $\{x_n\}$，$\{y_n\}$，我们定义

$$\rho_X(x,y)=\lim_{n\to\infty}\rho_{X_0}(x_n,y_n). \tag{7}$$

上式右方 $\rho_{X_0}(x_n,y_n)$ 是有意义的，因为 $x_n$，$y_n$ 都是 $X_0$ 之元，这个极限是否存在？注意 $\rho_{X_0}(x_n,y_n)$ 是实数，而实数系的完备性是证明了的，所以我们只需证明，$\{r_n\}=\{\rho_{X_0}(x_n,y_n)\}$ 是 **R** 中的柯西序列即可. 这是容易的. 由三角形不等式

$$\rho_{X_0}(x_n,y_n)\leqslant\rho_{X_0}(x_n,x_m)+\rho_{X_0}(x_m,y_m)+\rho_{X_0}(y_m,y_n),$$

故

$$\begin{aligned}r_n-r_m&=\rho_{X_0}(x_n,y_n)-\rho_{X_0}(x_m,y_m)\\&\leqslant\rho_{X_0}(x_n,x_m)+\rho_{X_0}(y_n,y_m).\end{aligned}$$

交换 $n$ 与 $m$ 又有

$$r_m-r_n\leqslant\rho_{X_0}(x_n,x_m)+\rho_{X_0}(y_n,y_m).$$

总之有

$$|r_n-r_m|\leqslant\rho_{X_0}(x_n,x_m)+\rho_{X_0}(y_n,y_m).$$

但 $\{x_n\}$，$\{y_n\}$ 都是 $X_0$ 中的柯西序列，所以当 $n,m\to\infty$ 时，$\rho_{X_0}(x_n,x_m)\to0$，$\rho_{X_0}(y_n,y_m)\to0$，而(7)之极限存在.

由(7)定义的 $\rho_X(x,y)$ 之值与 $x,y$ 之代表元的选法无关. 如果我们取 $x,y$ 的另外的代表元 $\{x'_m\}$ 与 $\{y'_m\}$，则因

$$\rho_{X_0}(x'_n,y'_n)\leqslant\rho_{X_0}(x_n,x'_n)+\rho_{X_0}(x_n,y_n)+\rho_{X_0}(y_n,y'_n),$$

$$\rho_{X_0}(x_n,y_n)\leqslant\rho_{X_0}(x_n,x'_n)+\rho_{X_0}(x'_n,y'_n)+\rho_{X_0}(y_n,y'_n)$$

有

$$|\rho_{X_0}(x_n,y_n)-\rho_{X_0}(x'_n,y'_n)|\leqslant\rho_{X_0}(x_n,x'_n)+\rho_{X_0}(y_n,y'_n)\to0.$$

所以(7)式定义的 $\rho_X(x,y)$ 确实是只与 $X$ 之元 $x,y$ 有关. 今证它适合度量所应适合的条件. 非负性以及对称性是明显的，所以下面只证明三角形不等式. 取 $x,y,z\in X$，它们以 $\{x_n\}$，$\{y_n\}$，$\{z_n\}$ 为代表元，于是由 $X_0$ 中的三角形不等式，有

$$\rho_{X_0}(x_n,z_n)\leqslant\rho_{X_0}(x_n,y_n)+\rho_{X_0}(y_n,z_n).$$

令 $n\to\infty$ 求极限即得 $X$ 中的三角形不等式

$$\rho_X(x,z)\leqslant\rho_X(x,y)+\rho_X(y,z).$$

$X_0$ 作为 $X$ 之子集的含义已在上面说明，今证 $x,y\in X_0\subset X$ 在 $X_0$ 中的距离与它们在 $X$ 中的距离相同. 把 $x,y$ 看作 $X$ 中之元来计算其距离时，应该取 $x,y$ 之任一代表元按(7)来计算. 但是

$x, y \in X_0$，其特征就是 $x$ 作为柯西序列的等价类时，可以用常驻序列 $\{x\}$ 与 $\{y\}$ 作为其代表元．于是由(7)有

$$\rho_X(x, y) = \lim_{n\to\infty}\rho_{X_0}(x_n, y_n).$$

但是当 $x, y \in X_0$ 时，我们规定了是用常驻序列作为其代表元的，所以可取 $\{x_n\} = \{x, \cdots, x, \cdots\}$，$\{y_n\} = \{y, \cdots, y, \cdots\}$．因此 $\rho_{X_0}(x_n, y_n) = \rho_{X_0}(x, y)$，而有 $\rho_X(x, y) = \rho_{X_0}(x, y)$，而 $X_0$ 中的度量与它作为 $X$ 之子集由 $X$ 得到的度量一致．这就是说，$X_0$ 是 $X$ 的子空间．

下面我们来证明 $X$ 是完备的，取 $X$ 中一个柯西序列 $\{\tilde{x}_n\}$．每一个 $\tilde{x}$ 都是 $X_0$ 中一个柯西序列的等价类．所以各取 $\tilde{x}_n$ 之一代表元

$$\begin{aligned}
&\tilde{x}_1 = \{x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}, \cdots\},\\
&\tilde{x}_2 = \{x_1^{(2)}, x_2^{(2)}, \cdots, x_n^{(2)}, \cdots\},\\
&\cdots\cdots\cdots\cdots\\
&\tilde{x}_n = \{x_1^{(n)}, x_2^{(n)}, \cdots, x_n^{(n)}, \cdots\},
\end{aligned}\tag{8}$$

每一个横行都是 $X_0$ 中的一个柯西序列，所以对每个固定的上标 $n$，$\lim\limits_{k,l\to\infty}\rho_{X_0}(x_k^{(n)}, x_l^{(n)}) = 0$，因此可以选一个 $k_n$ 使

$$\rho_{X_0}(x_m^{(n)}, x_{k_n}^{(n)}) \leqslant \frac{1}{n}, m > k_n.\tag{9}$$

在(8)中按上述 $k_n$ 的选取方法构造一个序列：

$$\tilde{x} = \{x_{k_1}^{(1)}, x_{k_2}^{(2)}, \cdots, x_{k_n}^{(n)}, \cdots\},\tag{10}$$

我们称(10)为(8)的对角线序列．今证 $\tilde{x}$ 是 $X_0$ 中的一个柯西序列，因此代表 $X$ 中的一个元，记作 $\tilde{x}$．为此以 $\tilde{x}_{k_n}^{(n)}$ 代表 $x_{k_n}^{(n)}$ 所成的常驻序列：

$$\tilde{x}_{k_n}^{(n)} = \{x_{k_n}^{(n)}, x_{k_n}^{(n)}, \cdots, x_{k_n}^{(n)}, \cdots\}.$$

于是，由(9)式知当 $k$ 充分大时

$$\lim_{m\to\infty}\rho(x_m^{(n)}, x_{k_n}^{(n)}) = \rho_{X_0}(\tilde{x}_n, x_{k_n}^{(n)}) \leqslant \frac{1}{n}.\tag{11}$$

所以，

$$\begin{aligned}
\rho_{X_0}(x_{k_n}^{(n)}, x_{k_m}^{(m)}) &= \rho_X(\tilde{x}_{k_n}^{(n)}, \tilde{x}_{k_m}^{(m)})\\
&\leqslant \rho_X(\tilde{x}_n, \tilde{x}_{k_n}^{(n)}) + \rho_X(\tilde{x}_n, \tilde{x}_m) + \rho_X(\tilde{x}_m, \tilde{x}_{k_n}^{(m)})\\
&\leqslant \rho_X(\tilde{x}_n, \tilde{x}_m) + \frac{1}{n} + \frac{1}{m}.
\end{aligned}$$

我们选 $\tilde{x}_n$ 正是为了使 $\{\tilde{x}_n\}$ 成为 $X$ 中的柯西序列，所以由上式，当 $n, m > N(\varepsilon)$ 时

$$\rho_{X_0}(x_{k_n}^{(n)}, x_{k_m}^{(m)}) < \varepsilon.$$

这样得知(10)所定义的 $\tilde{x}$ 确实是 $X_0$ 中的一个柯西序列，因此它定义 $X$ 中一个元素 $\tilde{x}$．我们现在

证明

$$\lim_{n\to\infty}\widetilde{x}_n=\widetilde{x}. \tag{12}$$

事实上，

$$\rho_X(\widetilde{x},\widetilde{x}_n)\leqslant\rho_X(\widetilde{x},\widetilde{x}_{k_n}^{(n)})+\rho_X(\widetilde{x}_n,\widetilde{x}_{k_n}^{(n)})\leqslant\frac{1}{n}+\rho_X(\widetilde{x},\widetilde{x}_{k_n}^{(n)}),$$

$\widetilde{x}_{k_n}^{(n)}$是常驻序列，$\widetilde{x}=\{x_{k_1}^{(1)},\cdots,x_{k_p}^{(p)},\cdots\}$，$p\to+\infty$，所以上式右方又可估计如下：

$$\leqslant\frac{1}{n}+\lim_{p\to\infty}\rho_{X_0}(x_{k_p}^{(p)},x_{k_n}^{(n)})<\varepsilon,$$

只要 $n$ 充分大，因此(12)式成立，而知 $X$ 是一个完备空间.

再证明 $X_0$ 在 $X$ 中处处稠密.事实上，在 $X$ 中取一个元$\widetilde{x}$，我们要在 $X_0$ 中找一个元素序列逼近它. $\widetilde{x}$ 既是$X$之元，当然就是一个 $X_0$ 柯西序列之等价类.取此等价类的一个代表元$\{x_1,x_2,\cdots,x_n,\cdots\}$，则有 $x_n\in X_0$，且 $\rho_{X_0}(x_n,x_m)<\varepsilon$ 只要$n,m>N(\varepsilon)$.再取一个常驻序列 $\widetilde{x}_n=\{x_n,x_n,\cdots,x_n,\cdots\}\in X_0\cap X$ 这里$n$ 固定，则 $\rho_X(\widetilde{x}_n,\widetilde{x})=\lim\limits_{m\to\infty}\rho_{X_0}(x_n,x_m)$，注意 $n$ 是固定的.如果再令 $n\to\infty$即有$\lim\limits_{n\to\infty}\rho_X(\widetilde{x}_n,\widetilde{x})=0$.这样得知 $X_0=X_0\cap X$ 在$X$ 中处处稠密.

最后证明 $X$ 在等距意义下是唯一的.实际上，若有另一个完备空间 $\overline{X}$，适合定理的要求，我们先证明 $X$ 与$\overline{X}$ 作为集合是一一对应的.为此，取任一 $y\in\overline{X}$，因为定理要求 $X_0$ 在 $\overline{X}$ 中稠密，所以一定有 $X_0$ 中的一个序列$\{y_1,y_2,\cdots,y_n,\cdots\}$适合 $\rho_{\overline{X}}(y_n,y)\to0$.这里 $y_n$ 既是$X_0$ 之元，$X_0$又是 $X$ 的子空间，且有度量：$\rho_{X_0}(\cdot,\cdot)=\rho_{\overline{X}}(\cdot,\cdot)$，所以由 $\rho_{\overline{X}}(y,y_n)\to0$ 知 $\rho_{\overline{X}}(y_m,y_n)=\rho_{X_0}(y_m,y_n)\to0$，即$\{y_1,y_2,\cdots,y_n,\cdots\}$是 $X_0$ 的柯西序列，所以它定义 $X$ 中的一个元$x\in X$.这样，$\overline{X}$ 中每取一个元$y$，就相应在 $X$ 中有一个元$x$.

再在 $X$ 中任取一个元$x$，我们在 $\overline{X}$ 中求它的一个对应元$y$.事实上，既然 $x\in X$，而 $X$ 中之元又是$X_0$ 中的一个柯西序列，所以 $x=\{x_1,x_2,\cdots,x_n,\cdots\}$，这里 $x_n\in X_0$，$\rho_{X_0}(x_n,x_m)\to0$，但$X_0$ 已嵌入在 $\overline{X}$ 中，而且是等距的，所以

$$\rho_{\overline{X}}(x_n,x_m)=\rho_{X_0}(x_n,x_m)\to0.$$

就是说$\{x_1,x_2,\cdots,x_n,\cdots\}$作为 $\overline{X}$ 中的序列是柯西序列. 由于 $\overline{X}$ 已假设是完备的，所以一定存在一个 $y\in\overline{X}$ 使$\lim\limits_{n\to\infty}\rho_{\overline{X}}(y_n,y)=0$.这个 $y$ 就是$x$ 的对应元.而且这个 $y$ 还是$x_n$ 在$\overline{X}$ 中的极限.至此我们知道 $X$ 与$\overline{X}$ 作为集合是一一对应的.不但如此，它们还是等距的.因为若取 $x,x'\in X$，它们的代表元是$\{x_1,x_2,\cdots,x_n,\cdots\}$与$\{x_1',x_2',\cdots,x_n',\cdots\}$，它们按上面的作法在 $\overline{X}$ 中的对应元是$y$ 与$y'$，则

$$\begin{aligned}\rho_X(x,x')&=\lim_{n\to\infty}\rho_{X_0}(x_n,x_n')=\lim_{n\to\infty}\rho_{\overline{X}}(x_n,x_n')\\&=\rho_{\overline{X}}(y,y').\end{aligned}$$

等距性得证，定理证毕.

在第一节中我们已看到，一旦证明了 **R** 的完备性，微积分中的许多基本定理都随之得证．对于一般度量空间又如何？有一些结果自然是谈不上了．例如单调有界数列的收敛性．因为在一般度量空间中没有办法定义一个序列(而不是数列)的单调性．同样，有界集的上、下确界也不好定义．(但是有界性概念仍然有效：度量空间 $X$ 的子集 $A$ 如果能放在以某点 $x_0$ 为心半径 $\rho$ 充分大的球中：$A\subset B_\rho(x_0)$，就说 $A$ 是有界集．但是不能说 $x_0$ 是原点，因为度量空间中没有原点的概念，只有在线性空间中才有这个概念．)但有一个重要定理在一般度量空间中仍成立，这就是区间套定理，但现在我们称为

**定理 7**(闭球套定理) 设 $X$ 为一完备度量空间，$\{A_n\}$ 是闭集所成的序列，$A_1\supset A_2\supset\cdots$，且 $A_n\subset B_{r_n}(x_n)$，$r_n\to 0$，则必有唯一一点 $x\in\bigcap\limits_{n=1}^{\infty}A_n$．

**证** 先证唯一性．设有两个不同点 $x,y\in\bigcap\limits_{n=1}^{\infty}A_n$，$x\neq y$ 则 $\rho(x,y)=r>0$，现在 $r_n\to 0$，故从某个 $N(\varepsilon)$ 开始，$r_N<\dfrac{r}{2}$，但 $x,y\in\bigcap\limits_{n=1}^{\infty}A_n\subset A_N\subset B_{r_N}(x_n)$，所以

$$\rho(x,y)\leqslant\rho(x,x_n)+\rho(x_n,y)\leqslant 2r_N<r.$$

这与 $\rho(x,y)=r$ 矛盾．

再证这样的点 $x$ 存在．我们从 $A_n$ 中任取一点 $y_n$，则当 $m,n>N(\varepsilon)$ 时，$y_m,y_n\in A_N$，从而 $\rho(y_m,y_n)\leqslant 2r_N<\varepsilon$，所以 $\{y_n\}$ 是 $X$ 中的柯西序列．由 $X$ 之完备性，知道一定存在一点 $x\in X$，使 $\lim\limits_{n\to\infty}\rho(y_n,x)=0$．$x$ 必在 $\bigcap\limits_{n=1}^{\infty}A_n$ 中，若不然，必有某个 $n_0$ 使 $x\notin A_{n_0}$，但 $A_{n_0}$ 是闭集，从而 $A_{n_0}=\mathring{A}_{n_0}\cup\mathrm{bd}\,A_{n_0}$．$x\notin A_{n_0}$，必为 $A_{n_0}$ 之外点．因此必有一个以 $x$ 为心，某个正数 $r$ 为半径的小球 $B_r(x)\cap A_{n_0}=\varnothing$，但我们又可找到某个 $N>n_0$，使得一切从 $N$ 开始的 $y_n$ 全适合 $\rho(y_n,x)<r$，即 $y_n\in B_r(x)$．但是这时 $y_n\in A_n\subset A_N\subset A_{n_0}$．所以 $B_r(x)\cap A_{n_0}\neq\varnothing$，这是矛盾的．

关于度量空间完备性的讨论就到此为止．

度量空间完备性概念有一个在整个数学中都极为重要的应用，这就是压缩映像原理．这里我们讲的是一个完备度量空间 $X$ 以及一个映像 $T:X\to X$，如果 $T$ 适合以下条件，则称 $T$ 为一个压缩映像：对 $X$ 中任意元 $x,y$：

$$\rho(Tx,Ty)\leqslant c\rho(x,y),\quad 0\leqslant c<1. \tag{13}$$

我们要证明

**定理 8**(压缩映像原理) 设 $X$ 为一完备度量空间，$T:X\to X$ 是一个压缩映像，则必有唯一一点 $x\in X$，使 $Tx=x$，$x$ 称为 $T$ 的不动点．

**证** 证明就是我们熟知的逐步逼近法．任取一点 $x_0\in X$，称为零次近似．一般说来当然不会有 $Tx_0=x_0$，记 $x_1=Tx_0$，$x_2=Tx_1$，$\cdots$，以及一般地

$$x_k=Tx_{k-1},\ k=1,2,\cdots. \tag{14}$$

$x_k$ 就是 $k$ 次近似．今证 $\{x_k\}$ 有极限存在，而此极限即为不动点．实际上，$\{x_k\}$ 是一个柯西序列，因为若取 $m>n$，

$$\rho(x_m,x_n)\leqslant\rho(x_m,x_{m-1})+\rho(x_{m-1},x_{m-2})+\cdots+\rho(x_{n+1},x_n).$$

但是右方第一项可以估计如下：

$$\begin{aligned}\rho(x_m,x_{m-1})&=\rho(Tx_{m-1},Tx_{m-2})\leqslant c\rho(x_{m-1},x_{m-2})\\&\leqslant c^2\rho(x_{m-2},x_{m-3})\\&\leqslant\cdots\leqslant c^{m-1}\rho(x_1,x_0).\end{aligned}$$

对其余各项也作类似估计，即得

$$\begin{aligned}\rho(x_m,x_n)&\leqslant(c^{m-1}+c^{m-2}+\cdots+c^n)\rho(x_1,x_0)\\&=\frac{c^n(1-c^{m-n})}{1-c}\rho(x_1,x_0)\\&\leqslant Mc^n\rho(x_1,x_0),M=\frac{1}{1-c}.\end{aligned}$$

因为$0\leqslant c<1$，由上式即知$\{x_n\}$是一个柯西序列，所以它有一个极限$x$. 今证$x$即不动点：$Tx=x$. 实际上，任给$\varepsilon>0$，必有$N(\varepsilon)$在，使当$k>N(\varepsilon)$时，$\rho(x_k,x)<\frac{\varepsilon}{2c}$，现取$n>N(\varepsilon)+1$，于是，因为$n-1>N(\varepsilon)$，$n>N(\varepsilon)$，而有

$$\begin{aligned}\rho(x,Tx)&\leqslant\rho(x_n,Tx)+\rho(x,x_n)=\rho(x,x_n)+\rho(Tx_{n-1},Tx)\\&\leqslant\rho(x,x_n)+c\rho(x,x_{n-1})<\varepsilon,\end{aligned}$$

由$\varepsilon$之任意性即知$\rho(x,Tx)=0$，亦即$Tx=x$.

最后证明不动点的唯一性. 设有两个不动点$x$和$y$，则

$$\rho(x,y)=\rho(Tx,Ty)\leqslant c\rho(x,y),$$

故$\rho(x,y)=0$，即$x=y$. 这就是不动点的唯一性，定理证毕.

压缩映像原理在整个数学中应用极多. 例如常微分方程的柯西问题的唯一性用逐步逼近法去证明，这在通常的教材中都会讲到. 下面我们再证明隐函数的存在定理，可以看到应用这个原理应注意些什么.

设有$(x,y)\in\mathbf{R}^2$的函数$f(x,y)$，我们的目的是从方程$f(x,y)=0$，证明必有唯一的函数$y=\varphi(x)$，适合这个方程，以及$y_0=\varphi(x_0)$. 这里$f(x_0,y_0)=0$. 首先的问题是选一个完备的度量空间$X$，现在我们令$X=C([a,b])$，即在紧集$[a,b]$上连续的函数之空间. 这个空间是赋范空间，因为对$\varphi\in C([a,b])$，即在$[a,b]$上连续的函数$\varphi(x)$，我们均可令

$$\|\varphi\|=\sup_{[a,b]}|\varphi(x)|.\tag{15}$$

因为紧集$[a,b]$上的连续函数必可达到上确界，所以$\|\varphi\|<+\infty$是有意义的. $\|\varphi\|$适合范数之各个要求也是自明的. 有了这个范数，即可定义$C([a,b])$中两点$\varphi$与$\psi$之距离为

$$\rho(\varphi,\psi)=\|\varphi-\psi\|.$$

现在要问，这个空间是否完备的？若$\{\varphi_n\}$是此空间的一个柯西序列，则当$m,n\to+\infty$时，$|\varphi_m(x)-\varphi_n(x)|\leqslant\sup_{[a,b]}|\varphi_m(x)-\varphi(x)_n|=\|\varphi_m-\varphi_n\|<\varepsilon\to0$. 但这就是$\{\varphi_n(x)\}$在$[a,b]$上一致收敛，故其极限函数$\varphi(x)$仍在$[a,b]$上连续. 所以$C([a,b])$是一个完备的度量空间. 其次我们要去作一个压缩映像. 我们知道，方程$f[x,\varphi(x)]=0$等价于方程

$$\varphi(x)=\varphi(x)-\frac{1}{M}f[x,\varphi(x)],\tag{16}$$

这里 $M$ 待定，于是令

$$T\varphi=\varphi-\frac{1}{M}f(x,\varphi). \tag{17}$$

为了使它成为压缩的，注意

$$T\varphi_1-T\varphi_2=(\varphi_1-\varphi_2)-\frac{1}{M}[f(x,\varphi_1)-f(x,\varphi_2)].$$

因此如果 $f(x,y)$ 对 $y$ 为连续可微，则

$$|T\varphi_1-T\varphi_2|=\left|(\varphi_1-\varphi_2)-\frac{1}{M}f'_\varphi(x,\varphi_1+\theta(\varphi_1-\varphi_2))(\varphi_1-\varphi_2)\right|.$$

这里 $0\leqslant\theta\leqslant1$. 如果我们对 $f$ 加上如下条件：

$$f'_\varphi\geqslant m>0,$$

并取 $M=\sup f'_\varphi$，立即有

$$|T\varphi_1-T\varphi_2|\leqslant\left(1-\frac{m}{M}\right)|\varphi_1-\varphi_2|.$$

先在右侧取 $x\in[a,b]$时之上确界，再在左侧取 $x\in[a,b]$时之上确界，立即有

$$\|T\varphi_1-T\varphi_2\|\leqslant c\|\varphi_1-\varphi_2\|,$$

$$0<c=1-\frac{m}{M}<1.$$

而 $T$ 是一个压缩映像.

但这里有一个重要问题. 压缩映像法是一个逐步逼近的迭代过程：先任意取零次近似 $\varphi_0(x)$，但我们要求 $\varphi_0(x_0)=y_0$，然后以 $\varphi_0$ 代入(17)来计算 $\varphi_1(x)$等等. 但是这就需要考虑 $\varphi_0$ 是否可以“代得进去”(17)中的 $f(x,\varphi)$. 因为，设若 $f(x,y)$定义于$[a,b]\times[-C,C]$，而 $\varphi_0(x)$ 可能在某些 $x$ 点在$[-C,C]$之外取值，这样 $f(x,\varphi_0(x))$就没有意义了. 所以在求 $\varphi_1(x)$之前先要证明 $\varphi_0(x)$适合以下的估计式：$|\varphi_0(x)|\leqslant C$. 对于 $\varphi_1(x),\varphi_2(x),\cdots$也都需要作出类似估计. 这种估计是在问题是否有解之前就要作出的，因此称为先验(a priori)估计. 如何作先验估计时常是求解一大类问题最主要的困难，而且是一些数学分支最基本的技巧. 我们在这里当然无法涉及. 所以我们规定 $f(x,y)$在$[a,b]\times(-\infty,+\infty)$上有定义而且连续来绕过一这一困难.

作了这些说明以后，我们就可以把最终的结果陈述如下：

**隐函数的存在定理** 设 $f(x,y)$在$[a,b]\times(-\infty,+\infty)$上连续且对 $y$ 连续可微，$f'_y(x,y)$适合以下的估计：$0<m\leqslant f'_y(x,y)\leqslant M$，且 $f(x_0,y_0)=0$，这里 $x_0\in[a,b]$，则必有唯一 $y=\varphi(x)$在$[a,b]$上连续，适合关系式 $y_0=\varphi(x_0)$使得 $f(x,\varphi(x))=0$.

**注** 这里陈述的隐函数定理当然也适用于多个自变量、多个未知函数的情况. 但是它与第三章中所陈述的稍有不同. 这是因为我们想回避先验估计造成的. 当然，我们最终得到的结果也稍强一些，我们得到的隐函数定义在整个闭区间$[a,b]$上，而不只是在 $x_0$ 附近.

这里我们还要求各次近似 $\varphi_k(x)$都适合 $\varphi_k(x_0)=y_0$，因此我们所考虑的空间其实是 $C^*=\{\varphi(x),\varphi(x)在[a,b]上连续而且\ \varphi(x_0)=y_0\}$. 它仍然是一个完备的度量空间，这些地方我们都未深究. 我们最后只想强调一下，这里的压缩映像原理，只是一大类不动点定理最简单的情况. 在下一节我们还会回到这个问题. 隐函数定理也是一个很重大的数学课题，但本书中我们不能

再讲了.

**3. 希尔伯特空间和巴拿赫空间** 关于度量空间的完备性问题,我们已作了详细的讨论,现在再回到本节开始时的一些具体空间.它们除了具有度量之外,还有线性结构,而且其度量都是由范数或内积而来.(应该提到,有这样的度量空间,它也有线性结构,也有完备性,但是其度量并非由范数来确定的.在广义函数论中就常遇到这类空间,所以也是很重要的,但我们不会讲到.)这些空间又都是完备的,如 $C(I)$,$L^p$ 等等都是.所以下面我们再转到这些具体的空间.

完备化定理告诉我们,完备化的 $X$ 是通过子空间 $X_0$"求极限"而来,而 $X_0$ 则在 $X$ 中稠密.例如 $\mathbf{R}$ 是通过有理数系 $\mathbf{Q}$"求极限"得到的,$C([a,b])$之元可以通过多项式的 $C([a,b])$极限即一致收敛极限得出(魏尔斯特拉斯定理).$X_0$ 中的元即$[a,b]$上的多项式.在研究函数空间时,$X_0$ 中的元时常比 $X$ 中的元更容易处理.例如,我们前面已多次说过,$L^p$ 函数可以用连续函数或阶梯函数,甚至 $C_0^\infty$ 函数在 $L^p$ 范数下去逼近. 而且还专门讲了如何应用磨光算子去构造这些 $C_0^\infty$ 函数.但基本的结论一直没有证明.在讲勒贝格积分时,我们介绍了一个鲁金定理:设有$[a,b]$上的几乎处处有限的可测函数 $f(x)$,则对任意 $\delta>0$ 必可找到一个闭集 $F\subset[a,b]$,使 $m([a,b]\setminus F)<\delta$,而 $f(x)$ 在 $F$ 上连续.但这个定理也未证明. 而在证明上述 $X_0$ 在 $X$ 中的稠密性时,还必须用到它.(此外在讲到广义函数时,例如对 $\delta$ 函数,我们作出过许多 $\delta$ 序列,都是用很光滑的函数去逼近 $\delta(x)$,但是这时的逼近已不是在某一个赋范空间的范数意义下,而是在广义函数意义下.一般地说,一个广义函数总可以用 $C_0^\infty$ 函数在广义函数意义下逼近.这时我们也说 $C_0^\infty$ 函数在广义函数空间中稠密.当然这与度量空间的完备化定理是不同的).这样看来,我们在定义或研究一个函数空间时,时常用该空间是某一类很规则的函数在某种范数下的完备化来刻画它.可见,度量空间的完备化不但是重要的概念,也是常用的方法与技巧.

由于常见的空间 $C(I)$,$L^p$ 都是完备空间,所以我们给出下面的

**定义 11** 完备的赋范线性空间称为巴拿赫空间(Banach space),完备的内积空间称为希尔伯特空间(Hilbert space).

当然我们也会区分实的和复的巴拿赫空间和希尔伯特空间,于是我们看到,$C(I)$,$L^p$ 都是巴拿赫空间,而 $L^2$ 空间与索伯列夫空间 $H^k$ 则是希尔伯特空间.

在这些空间中,希尔伯特空间最为重要.因为它是最接近欧氏空间 $\mathbf{R}^n$ 的线性空间.

欧氏空间 $\mathbf{R}^n$ 有两个基本的几何量.即向量的长与两个向量的夹角.在任何赋范空间中,都可以定义向量之长,但在内积空间中则还可以定义两个向量的交角,因为现在有施瓦兹不等式(6). 对于实的内积空间,$\langle x,y\rangle$ 是实数,所以由(6)可知,一定可以找到一个实数 $\alpha$,$|\alpha|\leqslant 1$,使

$$\langle x,y\rangle=\alpha\|x\|\cdot\|y\|$$

而我们又可以找到一个角 $\theta$,使 $\alpha=\cos\theta$,于是上式就成了实内积空间中的余弦定理

$$\langle x,y\rangle=\|x\|\cdot\|y\|\cos\theta.$$

$\theta$ 可以解释为 $x$ 与 $y$ 的夹角,特别是 $\langle x,y\rangle=0$ 时 $\theta=\dfrac{\pi}{2}$,而我们说 $x$ 与 $y$ 正交,0 与任何向量都正交.

对于复内积空间,当然不一定有实的 $\theta$ 适合上式,但是我们仍规定:若$(x,y)=0$,就说 $x$ 与 $y$ 正交,0 与任何向量都正交.

$x$ 与 $y$ 正交就记作 $x \perp y$，不论它们是实或复.

在所有线性空间中都可以定义向量的线性相关与线性无关，并由此可以定义维数. 但是在线性代数教本中，我们只讲有限维空间，因此只有有限多个线性无关向量. 对于一般线性空间则可能有无穷多个线性无关向量. 所谓线性空间 $E$ 有无穷多个线性无关向量即指任意给出 $N$ 个线性无关的向量 $x_1, \cdots, x_N \in E$，一定还可找到第 $N+1$ 个向量 $x_{N+1}$ 与它们线性无关. 这时我们说 $E$ 是无限维的. 对于内积空间 $H$，若它是有限维——$n$ 维的实内积空间，它必是赋有欧几里得范数的 $\mathbf{R}^n$. 若是无限维的，则可能或者找到一组无穷多个线性无关的向量 $\{e_1, e_2, \cdots\}$ 使得任一向量 $x \in H$ 均可写为

$$x = \sum_{i=1}^{\infty} x_i e_i, \tag{18}$$

或者不论怎么选可数无穷多个 $\{e_1, e_2, \cdots\}$，总有某个 $x \in H$ 不能写成(18)式. 前一种情况下，我们说 $H$ 具有可数无穷维，后一种情况下我们说 $H$ 具有不可数无穷维. 有限维和可数无穷维的内积空间都称为可分的(在拓扑空间理论中对空间的可分性还另有定义，那个定义与这里说的是一致的，但是我们将不给证明). 由于在实用中可分内积空间已经够用了，所以下面讲到内积空间(亦称前希尔伯特空间(pre-Hilbert space))，以及希尔伯特空间，都限于讨论可分的情况.

在往下讨论前，我们要说一下无穷展开式(18)的意义. 它的部分和 $S_N = \sum_{i=1}^{N} x_i e_i \in H$. (18)式即指 $\lim_{N \to \infty} S_N = x$，这里的极限是在 $H$ 的范数意义下讲的：

$$\lim_{N \to \infty} \| S_N - x \| = 0. \tag{19}$$

在有限维情况下，(18)中的 $\{e_1, \cdots, e_n\}$ 称为 $H$ 的一个基底，在无限维情况下，这样的 $\{e_1, e_2, \cdots\}$ 也称为基底. 对于内积空间 $H$，我们可用正交性概念来构造一个由互相正交的向量所成的基底，它很像 $\mathbf{R}^n$ 中的直角坐标系. 为此，我们先在非 0 的内积空间中作一个所谓就范正交系(orthonormal system，简称 o. n. 系)下面的作法称为格拉姆－施密特(Gram-Schmidt)手续，其实在有限维的情况下，线性代数教本中都介绍过这种方法. 令 $\{x_1, \cdots, x_k, \cdots\}$ 为 $H$ 中的线性无关向量组(有限多个或可数多个向量)，因为已设 $H \neq 0$，至少可以找到一个非零向量 $x_1$，而 $\{x_1\}$ 就是 $H$ 中一个线性无关向量组. 令 $e_1 = x_1 / \| x_1 \|$，则 $\| e_1 \| = 1$，若此向量组 $\{x_1, \cdots, x_h, \cdots\}$ 中仅含一个元，则过程结束，我们得到 $H$ 中一个 o. n. 系. 若此向量组还有其它元，则取一个 $x_2$ $(x_2 \neq 0)$. 考虑 $x_2$ 与 $e_1$ 所张的子空间 $\{\alpha_1 e_1\}$ 距离的最小值，即求一个 $c_1$，使 $\| x_2 - c_1 e_1 \|^2 = \min_{\alpha_1} \| x_2 - \alpha_1 e_1 \|^2$，实际上

$$\begin{aligned} \| x_2 - \alpha_1 e_1 \|^2 &= (x_2 - \alpha_1 e_1, x_2 - \alpha_1 e_1) \\ &= \| x_2 \|^2 - \overline{\alpha}_1 (x_2, e_1) - \alpha_1 \overline{(x_2, e_1)} + \alpha_1 \overline{\alpha_1} \\ &= \| x_2 \|^2 - |(x_2, e_1)|^2 + [(x_2, e_1) - \alpha_1][\overline{(x_2, e_1)} - \overline{\alpha_1}] \\ &= \| x_2 \|^2 - |(x_2, e_1)|^2 + |(x_2, e_1) - \alpha_1|^2 \\ &\geqslant \| x_2 \|^2 - |(x_2, e_1)|^2. \end{aligned}$$

右方显然与 $\alpha_1$ 的选取无关. 所以 $\| x_2 - \alpha_1 e_1 \|^2$ 确有最小值而在 $\alpha_1 = (x_2, e_1)$ 时达到. 令

$$y_2 = x_2 - (x_2, e_1) e_1.$$

因为 $x_2$ 与 $e_1=x_1/\|x_1\|$ 线性无关，所以 $y_2\neq 0$. 令

$$e_2=y_2/\|y_2\|=[x_2-(x_2,e_1)e_1]/\|x_2-(x_2,e_1)e_1\|,$$

则 $\|e_2\|=1$，而且 $(e_2,e_1)=[(x_2,e_1)-(x_2,e_1)(e_1,e_1)]/\|y_2\|=[(x_2,e_1)-(x_2,e_1)]/\|y_2\|=0$，即 $e_1\perp e_2$. 这样，$\{e_1,e_2\}$成为 $H$ 中的一个 o.n. 系. 若$\{x_1,x_2\}$尚未穷竭$\{x_1,\cdots,x_N,\cdots\}$则仿此做下去. 假设已得出$\{e_1,\cdots,e_{k-1}\}$成为 $H$ 中的一个 o.n. 系，而$\{x_1,\cdots,x_N,\cdots\}$中仍有 $x_k$ 与它们线性无关，也就是说$\{x_1,\cdots,x_{k-1}\}$尚未穷竭$\{x_1,\cdots,x_N,\cdots\}$. 记$\{e_1,\cdots,e_{k-1}\}$所张的子空间为 $H_{k-1}=\left\{\sum_{i=1}^{k-1}\alpha_ie_i,\alpha_i\text{ 为复常数}\right\}$，则对于$\{x_1,\cdots,x_N,\cdots\}$中的非零的 $x_k$，我们要找 $k-1$ 个复常数 $c_i,i=1,\cdots,k-1$，使

$$\left\|x_k-\sum_{i=1}^{k-1}c_ie_i\right\|^2=\min_{\alpha_i}\left\|x_k-\sum_{i=1}^{k-1}\alpha_ie_i\right\|^2.$$

仿照上面的作法，有

$$\begin{aligned}\left\|x_k-\sum_{i=1}^{k-1}\alpha_ie_i\right\|^2&=\left(x_k-\sum_{i=1}^{k-1}\alpha_ie_i,x_k-\sum_{i=1}^{k-1}\alpha_ie_i\right)\\&=\|x_k\|^2-\sum_{i=1}^{k-1}\overline{\alpha_i}(x_k,e_i)-\sum_{i=1}^{k-1}\alpha_i\overline{(x_k,e_i)}+\sum_{i=1}^{k-1}\alpha_i\overline{\alpha_i}\\&=\|x_k\|^2-\sum_{i=1}^{k-1}|(x_k,e_i)|^2+\sum_{i=1}^{k-1}|(x_k,e_i)-\alpha_i|^2\\&\geqslant\|x_k\|^2-\sum_{i=1}^{k-1}|(x_k,e_i)|^2.\end{aligned}$$

所以，这个最小值是可以达到的，而只要取 $\alpha_i=(x_k,e_i)$ 即可达到这个最小值. 令 $y_k=x_k-\sum_{i=1}^{k-1}(x_k,e_i)e_i$，显然 $y_k\neq 0$，再令 $e_k=y_k/\|y_k\|$，则$\{e_1,\cdots,e_k\}$成为一个 o.n. 系.

如果做了有限步就穷竭了$\{x_1,\cdots,x_N,\cdots\}$（它可能原来就只有有限多个元），我们就得到有限多个向量$\{e_1,\cdots,e_k\}$所成的 o.n. 系. 如果不能穷竭，就可以得到由可数多个$\{e_1,\cdots,e_N,\cdots\}$所成的 o.n. 系. 很明显，o.n. 系必为线性无关系.

需要注意的是，$e_k$ 必定可以只用 $x_1,\cdots,x_k$ 来线性表示，而不会涉及 $x_{k+1}$ 及以后的元，反过来 $x_k$ 也一定可以用 $e_1,\cdots,e_k$ 的线性组合来表示，而用不到 $e_{k+1}$ 及以后的元.（这其实就是线性代数中的斯坦尼兹(Steinitz)代换定理. 本来这个定理就不一定只限于有限维空间中有效.）如果$\{x_1,\cdots,x_N,\cdots\}$是 $H$ 的基底，则$\{e_1,\cdots,e_N,\cdots\}$也一定是 $H$ 的基底. 但与有限维情况有些区别. 在有限维情况下，基底中向量的个数恰好就是空间维数，所以不论是$\{x_1,\cdots,x_k\}$也好，$\{e_1,\cdots,e_k\}$也好，只要 $k=$ 空间维数，我们就得到了一个基底. 只不过$\{e_1,\cdots,e_k\}$是 o.n. 基底. 但在无限维 $H$ 的情况，哪怕 $H$ 是可分的，因为$\{e_1,\cdots,e_N,\cdots\}$中有（可数）无穷多个元素，"去掉"几个，例如留下$\{e_3,e_4,\cdots,e_N,\cdots\}$仍是含可数无穷多元素的 o.n. 系，但是若前者是基底，则后者不可能仍是基底，尽管它所包含的线性无关向量仍然与空间的维数相同——同为可数无穷大. 但不论如何，只要$\{x_1,\cdots,x_N,\cdots\}$是原空间的基底，则用上法作出的$\{e_1,\cdots,e_N,\cdots\}$仍是 $H$ 的基底. 这里我们再重复一下，说$\{x_1,\cdots,x_N,\cdots\}$是 $H$ 的基底，即是说，对任意 $x\in H$，必可找到常数（实的

或是复的) $\{c_i\}$ 使 $x = \sum_{i=1}^{\infty} c_i x_i$，亦即(19)，或 $\lim_{N\to\infty} \| S_N - x \| = 0$. 这里 $S_N$ 是这个级数的部分和，不论如何，我们得到了

**定理 9** 任一可分希尔伯特空间 $H$ 均有 o.n.基底.

并不是 $H$ 之任意可数无穷多个互相正交的单位长向量 $\{e_1, \cdots, e_N, \cdots\}$，$(e_i, e_j) = \delta_{ij}$ 都是基底，要看(18)式是否成立.

(18)式何时会成立？我们要看一下格拉姆－施密特手续的几何含意.为什么当 $\alpha_i = (x, e_i)$ 时 $\left\| x - \sum_{i=1}^{k-1} \alpha_i e_i \right\|^2$ 会达到最小？从几何上看，它是图 6－2－1 三角形△$OAC$ 中 $AC$ 边之平方，只有当 $AC \perp OC$ 时，$AC$ 才会达到最小，就是说只有作 $x$ 到子空间 $H_{k-1}$(即图 6－2－1 上直线 $OC$)的正交投影时，才可以达到最小.上面作的图的含意其实超过简单的示意图.因为在内积空间中毕达哥拉斯定理——即勾股定理是成立的：设有两个向量 $y \perp z$，则 $\| y - z \|^2 = (y - z, y - z) = \| y \|^2 - (y, z) - (z, y) + \| z \|^2 = \| y \|^2 + \| z \|^2$.我们把这个结果用于△$AA'C$，$AA' \perp OC$，实际上，$AA' = y_k = x - \sum_{i=1}^{k-1} (x, e_i) e_i$，所以 $(y_k, e_j) = (x, e_j) - \sum_{i=1}^{k-1} (x, e_i)(e_i, e_j) = (x, e_j) - \sum_{i=1}^{k-1} (x, e_i)\delta_{ij} = 0$，这里 $j = 1, 2, \cdots, k-1$. $OC$ 代表 $\{e_1, \cdots, e_{k-1}\}$ 所张的子空间 $H_{k-1}$. $AA'$ 既正交于这空间基底中的一切向量，自然也正交于整个子空间，从而也正交于其一切向量，特别也就正交于 $A'C = \sum_{i=1}^{k-1} (\alpha_i - (x, e_i)) e_i$. 故由勾股定理有

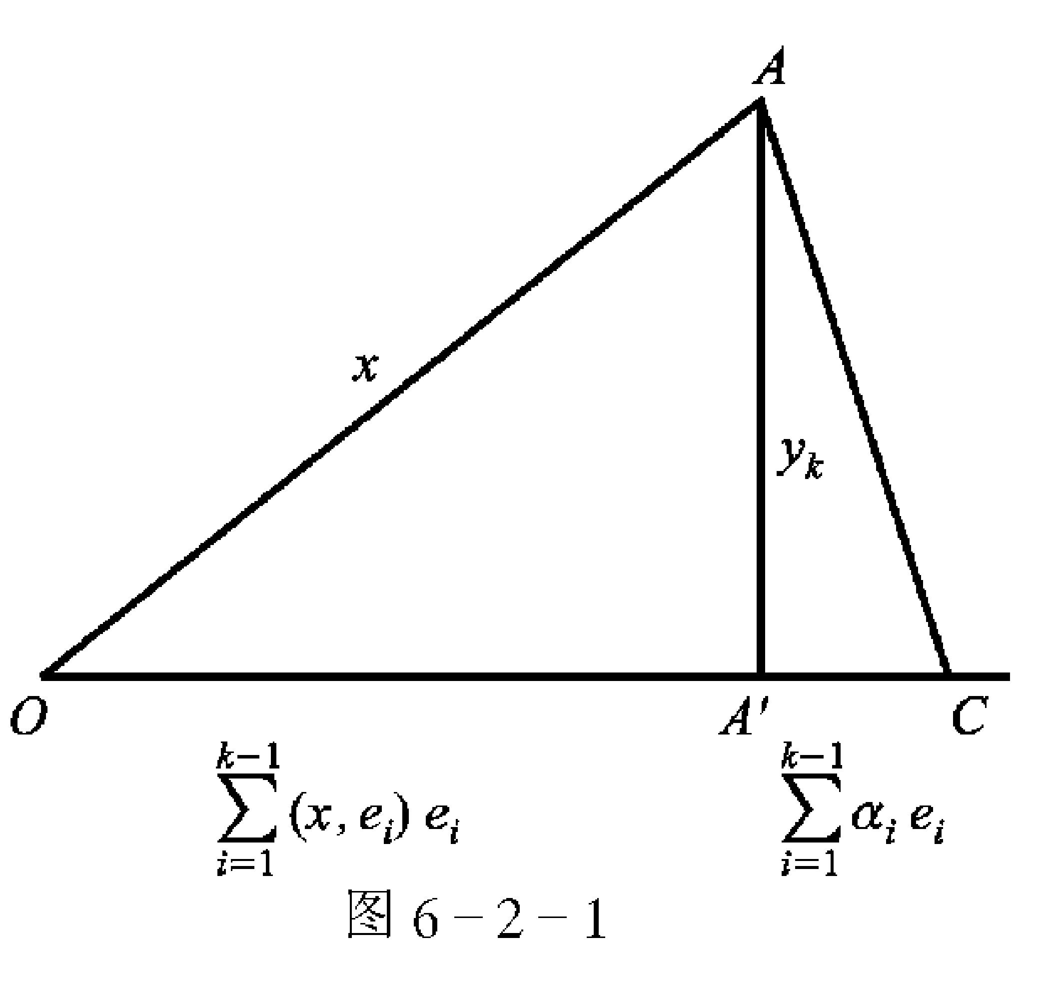


图 6－2－1

$$|AC|^2 = |AA'| + |A'C|^2.$$

但

$$|AC|^2 = \left\| x - \sum_{i=1}^{k-1} \alpha_i e_1 \right\|^2,$$

$$\begin{aligned}|AA'|^2 &= |OA|^2 - |OA'|^2 = \| x \|^2 - \left\| \sum_{i=1}^{k-1} (x, e_i) e_i \right\|^2 \\ &= \| x \|^2 - \left( \sum_{i=1}^{k-1} (x, e_i) e_i, \sum_{i=1}^{k-1} (x, e_i) e_i \right) \\ &= \| x \|^2 - \sum_{i,j} (x, e_i)\overline{(x, e_j)}\delta_{ij} \\ &= \| x \|^2 - \sum_{i=1}^{k-1} |(x, e_i)|^2,\end{aligned}$$

$$|A'C|^2 = \sum_{i=1}^{k-1} \| (\alpha_i - (x, e_i)) e_i \|^2 = \sum_{i=1}^{k-1} |\alpha_i - (x, e_i)|^2. \tag{20}$$

所以只有当 $A'$ 与 $C$ 重合，即 $AC \perp OC$ 时，$\| A'C \| = 0$，这时 $AC$ 才达到最小. 即是说当 $\alpha_i =$

$(x,e_i)$时，$\left\| x-\sum\limits_{i=1}^{k-1}\alpha_i e_i \right\| \to \min.$

如果(18)式成立，亦即(19)式成立，注意到 $S_N$ 就是这里的 $\sum\limits_i \alpha_i e_i$，就是此图上的 $OC$，因此(19)式左方之差即这里的 $|AC|^2$. 如果(19)式能成立，必有 $|AA'|^2\to 0$，就是说，向量 $x$ 的投影当 $k\to\infty$ 时，与原向量之"缺口 $AA'$ 会封闭起来. 这样我们就会得到判别一个 o.n. 系 $\{e_1,\cdots,e_N,\cdots\}$ 是否 $H$ 的基底的几何方法.

仿照上面的几何方法来计算. 我们有

**定理 10**(**贝塞尔不等式与帕塞瓦尔不等式**)　设 $H$ 为一可分的希尔伯特空间，$\{e_1,\cdots,e_N,\cdots\}$ 是其中的一个 o.n. 系(我们也容许此系只包含有限多个向量的特殊情况)，对任一向量 $x$，必有贝塞尔不等式(Bessel inequality)

$$\sum_{i=1}^{\infty}|(x,e_i)|^2\leqslant\|x\|^2, \tag{21}$$

成立. o.n. 系 $\{e_1,\cdots,e_N,\cdots\}$ 是 $H$ 的基底的充分必要条件是上式变成等式，称为帕塞瓦尔等式(Parseval's identily)或称封闭性方程式：

$$\sum_{i=1}^{\infty}|(x,e_i)|^2=\|x\|^2. \tag{22}$$

这时，我们说 $\{e_1,\cdots,e_N,\cdots\}$ 是一个完全的 o.n. 系

**证**　令 $x_i=(x,e_i)$，则仿照上面的证明有

$$\left\| x-\sum_{i=1}^{k-1}x_i e_i\right\|^2=\|x\|^2-\sum_{i=1}^{k-1}|x_i|^2.$$

因为上式左方非负，故有

$$\sum_{i=1}^{k-1}|x_i|^2\leqslant\|x\|^2,$$

令 $k\to\infty$ 即得(21)式.

再证帕塞瓦尔等式的必要性，若 $\{e_1,\cdots,e_N,\cdots\}$ 是 $H$ 之基底，则一定存在一列复数 $c_i$，使 $\left\| x-\sum\limits_{i=1}^{N}c_i e_i\right\|^2\to 0$，令 $x_i=(x_i,e_i)$，仿上面的作法有

$$\begin{aligned}\left\| x-\sum_{i=1}^{N}c_i e_i\right\|^2&=\|x\|^2-\left\|\sum_{i=1}^{k-1}x_i e_i\right\|^2+\sum_{i=1}^{N}|x_i-c_i|^2\\&\geqslant\|x\|^2-\sum_{i=1}^{N}|x_i|^2\geqslant 0\qquad\text{(贝塞尔不等式)}.\end{aligned}$$

若左方当 $N\to\infty$ 时趋于 0，自然有帕塞瓦尔等式成立.

充分性证明如下. 若(22)成立，则

$$\lim_{N\to\infty}\left\| x-\sum_{i=1}^{N}x_i e_i\right\|\to 0,$$

这就是说，一定能找到复数 $c_i$，使 $\left\| x-\sum\limits_{i=1}^{N}c_i e_i\right\|^2\to 0$ (取 $c_i=(x,e_i)$ 即可)，所以 $\{e_1,\cdots,e_N,\cdots\}$ 是基底. 证毕.

希尔伯特空间 $H$ 中的 o.n. 系 $\{e_,\cdots,e_N,\cdots\}$ 是一个基底，也就是说它是一个完全的 o.n. 系，还有另一个几何解释，即如果 $\{e_1,\cdots,e_N,\cdots\}$ 是 $H$ 中的一个完全的 o.n. 系，则不可能再找到一个非 0 的 $e\in H$，使 $e\perp e_i=0$，$i=1,2,\cdots,N,\cdots$. 事实上，若有 $e\in H$，$\|e\|=1$，而且 $e\perp e_i$，$i=1,2,\cdots$，则对 $e$ 应用(22)式当有

$$\|e\|^2=\sum_{i=1}^{\infty}|(e,e_i)|^2=0,$$

这与 $\|e\|=1$ 矛盾. 上面的论断之逆也是对的，即是说：如果 $\{e_i,\cdots,e_N,\cdots\}$ 不是 $H$ 中完全的 o.n. 系，则一定有一个 $x\in H$，使(21)成为严格的不等式：

$$\sum_{i=1}^{\infty}|(x,e_i)|^2<\|x\|^2.$$

但若令

$$y=\sum_{i=1}^{\infty}(x,e_i)e_i,$$

这个级数在范数意义下是收敛的. 因为考虑 $\sum_{i=1}^{\infty}(x,e_i)e_i$ 之两个部分和 $S_{N_1}$ 与 $S_{N_2}$，$N_2>N_1$，则容易证明，

$$\|S_{N_2}-S_{N_1}\|^2=\sum_{i=N_1+1}^{N_2}|(x,e_i)|^2.$$

由贝塞尔不等式，上式右方当 $N_1$ 和 $N_2\to\infty$ 时必趋于 0，所以 $\{S_N\}$ 是 $H$ 中的柯西序列，因此有一个 $y\in H$ 使 $y=\sum_{i=1}^{\infty}(x,e_i)e_i$. 显然

$$\|y\|^2=\lim_{N\to\infty}(S_N,S_N)=\sum_{i=1}^{\infty}|(x,e_i)|^2.$$

于是 $y\neq x$，因为由 $\{e_1,\cdots,e_N,\cdots\}$ 不完全的假设我们已知上式右方 $<\|x\|^2$. 但是我们可以证明 $(y,e_i)=(x,e_i)$. 事实上，若 $N$ 充分大($N>i$)，$(S_N,e_i)=\sum_{j=1}^{N}(x,e_j)(e_j,e_i)=\sum_{j=1}^{N}(x,e_j)\delta_{ij}=(x,e_i)$，所以 $(y,e_i)=(y-S_N,e_i)+(S_N,e_i)=(y-S_N,e_i)+(x,e_i)$. 但是 $|(y-S_N,e_i)|\leqslant\|y-S_N\|\|e_i\|=\|y-S_N\|\to0$. 所以 $(y,e_i)=(x,e_i)$ 对一切 $i$ 成立. 即是说 $(x-y)\perp e_i$，对一切 $i$ 成立. 令 $e=x-y$. 我们找到了 $H$ 中的一个非 0 的，正交于一切 $e_i$ 的元 $e$.

现在再回到 $L^2$ 空间. 第四章中. 我们着重指出，对于两个 $L^2([0,2\pi])$ 函数 $f(x)$ 与 $g(x)$，定义

$$(f,g)=\int_0^{2\pi}f(x)\overline{g}(x)\mathrm{d}x \tag{23}$$

成为一个内积. 在这样的内积下，$L^2([0,2\pi])$ 是一个复内积空间. 而且因为 $L^2(0,2\pi)$ 是一个完备空间(里斯－费希尔定理)，所以这是一个希尔伯特空间. 它是一个可分希尔伯特空间，因为它具有一个完全的 o.n. 系 $\left\{\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}nx}\right\}$，$n=0,\pm1,\pm2,\cdots$. 对于其中的任一元 $f$，

$$(f,e_n)=\frac{1}{\sqrt{2\pi}}\int_0^{2\pi}f(x)\mathrm{e}^{-\mathrm{i}nx}\mathrm{d}x$$

除了相差一个常数因子外就是 $f(x)$ 的傅里叶系数，而对 $L^2([0,2\pi])$ 函数，公式

$$\int_0^{2\pi}|f(x)|^2\mathrm{d}x=\sum_{n=0,\pm1,\cdots}|(f,e_n)|^2$$

是已经证明了的.

重要的事实是，$L^2([0,2\pi])$ 是一个"典型的"可分希尔伯特空间. 事实上，任给一个可分希尔伯特空间 $H$，定理 9 告诉我们，它一定具有一个完全的 o. n. 系 $\{e_i\}$，$i=0,\pm1,\cdots,\pm n,\cdots$. 这里我们对其指标 $i$ 重新作了编号，使它可取一切正、负整数值. 我们现在在 $H$ 与 $L^2([0,2\pi])$ 中建立如下的一一对应：对 $\tilde{f}\in H$，我们可以作函数

$$f(x)=\sum_j f_j\cdot\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}jx}$$

与 $\tilde{f}$ 相对应. 这里 $f_j=(\tilde{f},e_j)$. 这个函数确实是存在的，因为

$$\tilde{f}=\sum_j f_je_j,\tag{24}$$

而由于 $H$ 是一个希尔伯特空间，由(21)式，$\sum_j|f_j|^2<+\infty$. 因此这个级数在 $L^2([0,2\pi])$ 中收敛. 所以可以记其和为 $f(x)$. $f\in L^2([0,2\pi])$ 而且

$$f_j=\int_0^{2\pi}f(x)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\mathrm{i}jx}\mathrm{d}x.$$

反之，若有一个 $f(x)\in L^2([0,2\pi])$，则可用上式作其傅里叶系数 $f_j$，而用(24)式作一个 $\tilde{f}$，很容易证明 $\tilde{f}\in H$. 总之，$L^2([0,2\pi])$ 与 $H$ 有一个保持线性关系的同构. 在这个同构关系下，内积和范数都不变：设有另一个 $\tilde{g}\in H$，对应于另一个 $g(x)\in L^2([0,2\pi])$，它们的傅里叶系数是 $g_j=(\tilde{g},e_j)=\int_0^{2\pi}g(x)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\mathrm{i}jx}\mathrm{d}x$，则由帕塞瓦尔等式有

$$\|f\|^2=\sum_j|f_j|^2=\|\tilde{f}\|_N^2,$$

$$(f,g)_{L^2}=\sum_j f_j\tilde{g}_j=(\tilde{f},\tilde{g})_H.$$

所以这种对应保持内积与范数均不变. 总之是等距同构. 希尔伯特空间除了有线性结构以外，就只有内积与范数，决定了其所有性质. 在这个意义下，所有的可分希尔伯特空间，本质上都是 $L^2([0,2\pi])$.

关于度量空间以及赋范线性空间我们就讲到这里为止. 度量空间的另一些性质如紧性与连通性就都移到下节去讨论. 这一方面是由于本节篇幅太长，同时也只讨论度量空间，而紧性与连通性在一般的拓扑空间中讨论更合适，所以宁可在下一节讲. 在那里可以更清楚地看到，度量空间中的紧性与一般拓扑空间中的紧性有什么区别.

## §3 拓 扑 空 间

**1. 拓扑空间的基本概念** 度量空间是最直观的拓扑空间，它通过两点间的距离来定义开集、闭集、邻域、收敛性等等，与微积分中的讲法很接近. 而且在这个基础上我们特别建立了巴拿

赫空间和希尔伯特空间．那么为什么还要进一步抽象化呢？一方面固然是因为确有不可定义度量的空间(然而这类空间并不常见，有许多是属于"病态空间"之列)；另一方面，有许多情况下使用度量并不方便，这样就使得我们进一步考虑，在度量"后面"是否还有更深一层的结构？我们弄清楚了这些结构，就能更深刻地理解微积分学的一些概念、方法之实质．例如在§1中关于紧性的问题就有一个未证明的定理，讲到紧和列紧——即一个序列是否有收敛子序列的问题，又如在那里讲到连续函数的一些性质，我们只说关于连续函数可以取中间值的定理和其它两个性质(达到最大最小值以及一致连续性)是不相同的，它涉及连通性．那么什么是连通性呢？我们学过的微积分知识中还有哪一些是涉及连通性的呢？因此，我们的目的并不是要去研究一些很特殊的空间．虽然我们主张，如果一个空间很有用，就应该去讨论它，哪怕不太容易也要讨论．我们的目的在于探究一些更深刻的结构，以便用来解决更多的问题．

举一个例子，设函数 $y=f(x)$ 在 $x=x_0$ 点连续，其定义是任给一个 $\varepsilon>0$，必有相应的 $\delta(\varepsilon)>0$ 存在，使当 $|x-x_0|<\delta$ 时，$|y-f(x_0)|<\varepsilon$．如果用邻域的语言来陈述就是：设有 $f(x_0)$ 的一个邻域 $V$，必可找到 $x_0$ 的相应邻域 $U$，使得 $f(U)\subset V$．这里，$\varepsilon$ 和 $\delta$ 分别是用以度量 $V,U$ 的两个数．也就是说，邻近的程度是用数来刻画的．现在要问，如果不用数的大小可否仍然讨论"邻近"的概念？当然这样做会产生一些困难，例如 $f(x_0)$ 的邻域 $V$ 若用 $|y-f(x_0)|<\varepsilon$ 来刻画，则两个邻域可以"比较大小"，可以讨论一串邻域逐渐趋向一点等等，如果没有度量，这些概念如何刻画？把这一类问题暂时放在一边，应该说，"邻近"的概念是否一定要用度量来刻画是一个问题．

如果不用度量来刻画，就只有用公理化的方法来"定义"什么样的集合可以说是 $x_0$ 或 $f(x_0)$ 的邻域，例如至少应该要求 $x_0\in$ ($x_0$ 的任一邻域)……

认识到有必要这样来讨论空间中有关连续性的问题或称为拓扑问题，经历了一个漫长的时期．可以追溯到19世纪后半叶，其奠基者应该算是黎曼．到了20世纪初，人们才得到了共识，有了拓扑空间的理论．这个理论采取了公理化的形式，也是那个时代数学的总的潮流决定的．对于一个空间的拓扑性质，即有关连续性的性质，可以从一个或几个基本概念开始讲起，至于选取哪一个概念，有不同的作法．有主张从邻域开始，有主张从闭包开始，而后来大家公认的讲法则是从开集开始．这段历史我们不能细说了，只是要提醒，当年拓扑学的发展是为了更深刻地研究许多数学和物理学问题——一个重大问题是天体力学问题，而不只是为了对数学概念作更精确的描述．

下面我们就按拓扑学发展的这种精神，从最基本的概念开始作公理化的叙述．

**定义1** 空间(即一个集合) $X$ 之一个拓扑 $\tau$ 就是指 $X$ 的一族子集 $\mathscr{O}$，其中之元(都是 $X$ 的子集)称为开集，$\tau$ 应该适合以下的条件：

(i) $X$ 和 $\varnothing$ 都是开集：$X\in\tau$，$\varnothing\in\tau$；

(ii) 任意多个 $\tau$ 中之元，即任意多个开集之并仍为开集：$\bigcup\limits_{\lambda}U_\lambda\in\tau$ $(U_\lambda\in\tau)$；

(iii) 有限多个 $\tau$ 中之元，即有限多个开集 $U_1,\cdots,U_N$ 之交仍为开集，即 $\bigcap\limits_{k=1}^{N}U_k\in\tau$．$X$ 赋以这样的拓扑后称为一个拓扑空间．

上一节中我们看到在度量空间 $X$ 中，以球为基础，定义了开集．我们还特别约定空集是开

集.全空间是开集则很容易理解.球是一个与度量(半径长)有关的概念,所以那里的开集并非是先验地加以规定的,它是否有现在的定义中的(i)～(iii)的性质,是应该证明的.我们也确实证明了它们(§2 定理 2).这样我们看到,在度量空间中按§2 定义 4 所定义的开集,确实符合现在所给的开集的定义.于是读者会问,按现在的定义所给的开集,对于度量空间,是否恰好就是§2 定义 4 的开集?这是不一定的.因为§2 定义 4 如上所述是以度量为基础的,而这里的定义完全不涉及度量.从完全不涉及度量的定义出发,要能推导出基于度量的概念,是很不容易的.一定要对 $X$ 加上一些限制才行.这时,我们就说 $X$ 可度量化(metrizable).而且重要的是,同一个集合 $X$(我们暂时不用空间这个词)可以赋予完全不同的拓扑,构成完全不同的空间,怎么可能都是度量空间呢?大家都会看到,$X$ 上至少有以下两种拓扑:

i)　$(X,\tau_0)$($\tau_0$ 就是指适合定义 1 的(i)～(iii)的开集族或者称为一个拓扑),这里 $\tau_0=\{\varnothing,X\}$,它显然适合定义 1 之要求.$\tau_0$ 称为平凡拓扑.

ii)　$(X,\tau_1)$,$\tau_1=\{X$ 之所有子集$\}$.当然 $\varnothing$ 和 $X$ 都是 $X$ 的子集.$\tau_1$ 称为离散拓扑.

iii)　如果可能在 $X$ 上引入度量 $\rho$,则可以用这个度量来定义§2 定义 4 那样的拓扑,因此,这种拓扑称为度量拓扑.

于是我们看到,任给一个集合 $X$,都至少可以赋以 $\tau_0$ 与 $\tau_1$ 两种拓扑.但判断一个集合是否可以赋以度量拓扑则是很困难的事.这样我们也就看到,度量空间只是一种很特殊的拓扑空间.因为同一个集合 $X$ 可以赋予不同的拓扑而成不同的拓扑空间,所以有时我们也把有关的拓扑 $\tau$ 写出来而记一个拓扑空间为 $(X,\tau)$.

同一集合 $X$ 可以赋予不同的拓扑,这就产生了拓扑的比较问题.设同一集合 $X$ 上有两个拓扑 $\tau$ 与 $\tau'$,即指定两个子集族满足定义 1 的各项要求,并分别称其中之元为 $\tau$ 开集与 $\tau'$ 开集.若一切 $\tau$ 开集均为 $\tau'$ 开集,即由 $U\in\tau$ 可得 $U\in\tau'$,就说 $\tau$ 弱于(粗于)$\tau'$,同时也就说 $\tau'$ 强于(细于)$\tau$.所以,平凡拓扑弱于一切拓扑,而离散拓扑则强于一切拓扑.但是并非任意两个拓扑都可以比较强弱.

**定义 2**　设 $(X,\tau)$ 为一拓扑空间.$U\subset X$ 称为这个拓扑空间(或简单就说是"这个拓扑")的闭集,如果其余集 $U^c=X\setminus U$ 是一个开集.

闭集有下面的性质(证明很容易,故略去):

**定理 1**　若 $(X,\tau)$ 之闭集族记为 $\mathscr{F}$,则

(i) $X,\varnothing\in\mathscr{F}$;

(ii) 有限多个闭集之并仍为闭集,即若 $U_1,\cdots,U_N\in\mathscr{F}$,则 $\bigcup\limits_{\lambda=1}^{N}U_\lambda\in\mathscr{F}$;

(iii) 任意多个闭集之交仍为闭集:即若 $U_\lambda\in\mathscr{F}$,则 $\bigcap\limits_{\lambda}U_\lambda\in\mathscr{F}$.

**定义 3**　设 $(X,\tau)$ 是一个拓扑空间,$x\in X$ 的邻域是指适合以下条件的 $X$ 之子集 $\Omega(x)$:$x\in\Omega(x)$ 而且存在一个开集 $U$:$x\in U\subset\Omega(x)$.

这个定义与§2 定义 5(度量空间中邻域的定义)是一致的,$x$ 之邻域的集合称为 $x$ 之邻域系,记作 $\mathscr{N}_x$.

**定理 2**　设 $(X,\tau)$ 是一个拓扑空间,$\mathscr{N}_x$ 为 $x\in X$ 之邻域系,则

(i) 若 $U_\lambda\in\mathscr{N}_x$,则 $\bigcup\limits_{\lambda}U_\lambda\in\mathscr{N}_x$;

(ii) 若 $U_1,\cdots,U_N\in\mathcal{N}_x$，则 $\bigcap_{k=1}^{N}U_k\in\mathcal{N}_x$；

(iii) 若 $U\in\mathcal{N}_x$，则 $x$ 必有一个邻域 $\widetilde{U}$，使 $U$ 为 $\widetilde{U}$ 中任一点的邻域.

(iv) 若 $U\in\mathcal{N}_x$，$V\supset U$，则 $V\in\mathcal{N}_x$；

定理 1，2 之证明均略去，因为它们与度量空间中的相应定理（§2 定理 3，定理 4）完全一样. 开邻域的概念，以及关于 $X$ 和 $\varnothing$ 既为开集又为闭集的讨论在这里也适用.

关于开集与邻域的关系，很容易证明

**定理 3** 设 $(X,\tau)$ 是一个拓扑空间，$A\subset X$ 非空，$A$ 为开集之充分必要条件是：$A$ 是其每一点的邻域.

**证 必要性** 是明显的，因为定义 3 中的 $\Omega(x)$ 与 $U$ 可以都选为 $A$.

**充分性** 设对任一点 $x\in A$，$A$ 均为其邻域，由定义 3 必存在开集 $U_x$，使 $x\in U_x\subset A$，取 $A$ 中所有的点 $x$，并取 $U_x$ 之并，于是 $\{x\}\subset\bigcup_x U_x\subset A$，亦即 $A\subset\bigcup_x U_x\subset A$，所以 $A=\bigcup_x U_x$. 由开集的性质 (ii)，$A$ 是开集.

下面我们把度量空间中集合 $A$ 的内域、闭包等概念也都移到拓扑空间 $(X,\tau)$ 中来. 这里的结果与 §2 的定义 6、7 和定理 5 也都是一样的，所以证明全都略去.

**定义 4** 设 $(X,\tau)$ 为一拓扑空间，$A\subset X$，$x\in X$.

(i) $x$ 称为 $A$ 之内点，如果 $x$ 有一个开邻域 $U\subset A$；

(ii) $x$ 称为 $A$ 之外点，如果 $x$ 是 $A$ 之余集 $A^c$ 之内点；

(iii) 若 $x$ 既非 $A$ 之内点，又非 $A$ 之外点，则称 $x$ 是 $A$ 的边界点.

**定义 5** 同上，$A$ 之内点的集合称为 $A$ 之内域，记作 $\mathring{A}$ 或 $\mathrm{int}A$；$A^c$ 之内域的余集 $[\mathrm{int}(A^c)]^c$ 称为 $A$ 之闭包，记作 $\overline{A}$；边界点之集合记为 $\mathrm{bd}A=X-[\mathring{A}\cup(\mathrm{int}A^c)]$，称为 $A$ 的边界.

**定理 4** $\mathring{A}$ 是一切含于 $A$ 中的开集之并，所以是含于 $A$ 内的最大开集；$\overline{A}$ 是一切包含 $A$ 的闭集之交，所以是包含 $A$ 的最小闭集；$\mathrm{bd}A$ 也是闭集，而且

$$\overline{A}=\mathring{A}\cup\mathrm{bd}\,A=A\cup\mathrm{bd}\,A,$$

$$\mathring{A}=\overline{A}\setminus\mathrm{bd}\,A=A\setminus\mathrm{bd}\,A. \tag{1}$$

从这个定理还可以看到，$A$ 为开（闭）集的充分必要条件是 $A=\mathring{A}$ $(A=\overline{A})$.

我们还要引入一个很有用的概念——导集. 为此先给出以下的定义：

**定义 6** 设 $(X,\tau)$ 为一拓扑空间，$A\subset X$，$x\in X$ 称为 $A$ 之极限点，如果 $x$ 之任一邻域 $\Omega(x)$ 中均含有 $A$ 中除 $x$ 以外的其它点，极限点之集合称为 $A$ 之导集，记作 $A'$.

极限点的概念很明显是通常微积分教材中讲到的点的序列之极限点的概念的推广. 我们不妨看一下，$x\notin A'$ 是什么意思. 由定义 6，$x\notin A'$ 显然是指 $x$ 有某一邻域 $\Omega(x)$ 或者与 $A$ 不相交，即其中不含 $A$ 之点，这时 $x$ 是 $A$ 的外点；或者 $\Omega(x)$ 中只有一点 $x\in A$，这时 $x$ 是 $A$ 的孤立点，

所以极限点就是在 $X$ 中除去 $A$ 之外点以及 $A$ 之孤立点后余下的点．极限点也称为聚点．

我们在 §1，§2 中都讲到了处处稠密的概念，而且看见了它在微积分以至整个分析数学中的重要性．现在我们可以给它一个正式的定义．

**定义 7**　设 $(X,\tau)$ 是一拓扑空间，$X$ 之子集 $A$ 称为在 $X$ 中稠密，就是指 $\overline{A}=X$．

在微积分中我们经常会遇到这样的情况，即要在一个拓扑空间的某一子集 $A$ 上讨论一个数学问题．最简单的例子是讨论函数 $f(x_1,\cdots,x_n)$ 在某点 $x_0=(x_1^0,\cdots,x_n^0)$ 的连续性．如果这点是 $f$ 之定义域 $A$ 的内点，自然没有任何困难．如果 $x_0\in \mathrm{bd}A$，则条件 $\lim\limits_{x\to x_0} f(x)=f(x_0)$ 应该改成 $\lim\limits_{\substack{x\to x_0\\ x\in A}} f(x)=f(x_0)$．如果 $f(x)$ 是一元函数，而 $A$ 是一个区间 $[a,b]$，我们通常都说在其端点应该用左、右连续来代替连续．但若 $f(x_1,\cdots,x_n)$ 是多元函数，而 $A$ 是 $\mathbf{R}^n$ 的某一子集，则 $\mathrm{bd}A$ 的构造可以非常复杂．这时把 $A$ 本身就当作一个拓扑空间就方便多了．因此，我们宁可用 $(A,\tau')$ 代替 $(X,\tau)$．问题是 $\tau'$ 如何选择．我们当然希望 $\tau'$ 与 $\tau$ 有关．所以我们一定选用所谓子空间拓扑．

**定义 8**　设 $(X,\tau)$ 为一拓扑空间，而 $A\subset X$，在 $A$ 中定义拓扑如下：$U\subset A$ 为开集，当且仅当 $U=V\cap A$，而 $V$ 是 $(X,\tau)$ 中的开集．$A$ 上的这个拓扑称为子空间拓扑．若记它为 $\tau'$，我们用 $(A,\tau')\subset(X,\tau)$ 表示不仅 $A$ 是 $X$ 的子集合，而且 $\tau'$ 是子空间拓扑（又称为 $X$ 在 $A$ 上诱导的拓扑，诱导拓扑或相对拓扑）．

以下凡说到某拓扑空间是 $X$ 的子空间时，总是指其上赋有子空间拓扑．当然这里应该验证，$\tau'$ 确实是拓扑，即它适合定义 1 中的 (i) 到 (iii)，这是很容易的，所以留给读者自行验证．

这样一来，关于左、右连续，读者自然会看出，即是在子空间拓扑下的连续性．而且读者不妨证明一下一个简单的事实：若 $x_0$ 是 $A$ 的孤立点，则任一定义在 $A$ 上的函数 $f(x)$ 在 $x_0$ 处都连续．这个例子会帮助读者不必费很大的劲去专门讨论 $\mathrm{bd}A$ 的构造．

拓扑学的主题不但是拓扑空间，还有拓扑空间之间的映射．其中最重要的是两个拓扑空间 $(X,\tau)$，$(Y,\sigma)$ 之间的连续映射．

**定义 9**　$f:(X,\tau)\to(Y,\sigma)$ 称为连续的，如果 $(Y,\sigma)$ 之每个开集 $V$ 在 $f$ 下的原像 $f^{-1}(V)=U$ 是 $(X,\tau)$ 的开集．

但是我们在微积分教材中习惯的是一个映射（一个函数）在 $x_0$ 点的连续性，而 $f$ 在整个区域上为连续即指 $f$ 在每点上均为连续．在一般拓扑空间中我们也可以这样做，而有

**定理 5**　$f:(X,\tau)\to(Y,\sigma)$ 为连续，当且仅当 $f$ 在 $X$ 之每一点 $x_0$ 均为连续，即对 $f(x_0)$ 在 $(Y,\sigma)$ 中之每个邻域 $V_{f(x_0)}$ 均可找到 $x_0$ 在 $(X,\tau)$ 中的一个邻域 $U_{x_0}$，使 $f(U_{x_0})\subset V_{f(x_0)}$．

**证**　(i) **必要性**　设 $f$ 为连续，任取一点 $x_0$，并记 $f(x_0)=y_0$，任取 $y_0$ 在 $(Y,\sigma)$ 中的邻域 $V$，自然有 $f^{-1}(V)=U\subset X$，今证 $U$ 必为 $x_0$ 的邻域．首先，$x_0\in U$ 是明显的，其次 $V$ 既是 $y_0$ 之邻域，必有含 $y_0$ 的开集 $V_1\subset V$．显然 $x_0\in f^{-1}(V_1)\subset U$，但由假设，$f$ 是连续的，开集 $V_1$ 之原像 $f^{-1}(V_1)=U_1$ 是 $(X,\tau)$ 中的开集，而 $x_0\in U_1\subset U$．所以由邻域之定义，$U$ 是 $x_0$ 之邻域，而且 $f(U)=V$，故 $f$ 在每一点 $x_0$ 均为连续．

(ii) **充分性**　设 $f$ 在 $X$ 之每一点处均为连续，任取 $(Y,\sigma)$ 中的开集 $V$，今证 $f^{-1}(V)$ 是 $X$ 之开集．如果 $f^{-1}(V)=\varnothing$，$\varnothing$ 当然是 $(X,\tau)$ 中的开集．如果 $f^{-1}(V)\neq\varnothing$，则令 $x_0$ 为 $U=f^{-1}(V)$ 中任一点，$f(x_0)=y_0$ 即成为 $V$ 中之任意点．由邻域之性质，开集 $V$ 是其任意点之邻域，故由假

设应有 $x_0$ 之邻域 $U_{x_0}$ 使 $f(U_{x_0})\subset V$，从而 $U_{x_0}\subset U$，但 $U_{x_0}$ 中应有一个开集包含 $x_0$ 点，我们就用 $U_{x_0}$ 记这个开集，所以可设 $U_{x_0}$ 为开集. 如果 $x_0$ 遍取 $U$ 中各点，则

$$U=\bigcup_{x_0\in U}\{x_0\}\subset\bigcup_{x_0\in U}U_{x_0}\subset U.$$

所以 $U=\bigcup\limits_{x_0\in U}U_{x_0}$，右方是开集 $U_{x_0}$ 之并，所以也是开集. 即是说 $f^{-1}(V)=U$ 作为 $V$ 之原像是开集. 因此 $f$ 连续. 证毕.

读者应该注意，连续性的定义为“开集的原像仍是开集”，而不是“开集的像是开集”. 即是说，若 $U$ 是 $(X,\tau)$ 之开集，$f(U)$ 不一定是 $(Y,\sigma)$ 之开集. 如果一个映射 $f:(X,\tau)\to(Y,\sigma)$ 映 $X$ 中之一切开(闭)集为开(闭)集，即上述 $f(U)$ 为开(闭)集，这种映射称为开(闭)映射，开(闭)映射的概念在许多数学问题中很重要，但无论如何，它不必是连续映射. 我们不来举例了，只是再提醒一下，必须把开映射和连续映射区别开.

在许多数学分支中还会遇到以下的问题：已知 $f:X\to Y$，如果在 $X$ 或 $Y$ 中已赋有了拓扑，问在 $Y$ 或 $X$ 应该给怎样的拓扑才能使 $f$ 为连续？例如已给了拓扑空间 $(X,\tau)$，则对于 $U\in\tau$ (即对 $\tau$ 中的开集 $U$)，记 $f(U)=V$，我们有 $U=f^{-1}(V)$. 如果我们定义这样的 $V$ 就是 $Y$ 中一个开集，如果这能够构成一个拓扑 $\sigma$，则 $f:(X,\tau)\to(Y,\sigma)$ 显然是连续的. 因为 $\sigma$ 中的开集，就是按照 $\tau$ 中开集 $U$ 的像来定义的，亦即按照其原像为 $\tau$ 中的开集这一要求来定义的. 问题是这样的 $\sigma$ 是否为一拓扑. 显然，$\varnothing$ 与 $Y$ 都属于 $\sigma$，因为其原像为 $\varnothing$ 与 $X$ 都是 $\tau$ 中的开集. 若 $V_\lambda\in\sigma$ 对一切 $\lambda$ 成立，而且 $U_\lambda=f^{-1}(V_\lambda)$，则 $U_\lambda$ 为 $\tau$ 之开集，而 $\bigcup\limits_\lambda U_\lambda$ 也是 $\tau$ 之开集，但 $f(\bigcup\limits_\lambda U_\lambda)=\bigcup\limits_\lambda f(U_\lambda)=\bigcup\limits_\lambda V_\lambda$，所以 $\bigcup\limits_\lambda V_\lambda\in\sigma$. 此外，对有限多个 $V_k\in\sigma$，$k=1,2,\cdots,N$，令其原像为 $U_k=f^{-1}(V_k)$，则 $U_k\in\tau$，而 $\bigcap\limits_{k=1}^N U_k\in\tau$. 但是 $f(\bigcap\limits_{k=1}^N U_k)=\bigcap\limits_{k=1}^N f(U_k)=\bigcap\limits_{k=1}^N V_k$，所以 $\bigcap\limits_{k=1}^N V_k$，既然其原像是 $\tau$ 之开集，则其本身应是 $Y$ 中之开集，即 $\bigcap\limits_{k=1}^N V_k\in\sigma$. 所以这样作出的 $\sigma$ 确实是一个拓扑. 如果在这样的 $\sigma$ 之外再添加哪怕一个子集 $\widetilde{V}$ 作为开集，则 $\widetilde{V}$ 之原像 $\widetilde{U}$ 必不在 $\tau$ 中，因为 $\tau$ 中之子集都被 $f$ 映射到 $\sigma$ 内而不可能在 $\sigma$ 外. 所以，只要 $Y$ 中再加一个“开集”也会破坏 $f$ 的连续性. 所以我们说 $\sigma$ 是使 $f$ 连续最强(细)的拓扑.

与此相似，若在 $Y$ 中已有拓扑 $\sigma$，令 $V\in\sigma$，我们定义 $V$ 在 $f$ 下的原像 $U=f^{-1}(V)$ 为 $X$ 中的开集，和上面一样，这样定义的开集族 $\tau$ 确实构成一个拓扑 $\tau$，而且 $f:(X,\tau)\to(Y,\sigma)$ 是连续的，但若把 $\tau$ 中的开集减少一个 $\widetilde{U}=f^{-1}(\widetilde{V})$，则 $\sigma$ 开集 $\widetilde{V}$ 之原像 $\widetilde{U}$ 不再为开，因此 $f$ 不再连续. 这就是说，要保证 $f$ 为连续，$\tau$ 中的开集一个也不能少. 所以定义 $\sigma$ 开集之原像为 $\tau$ 开集，这样得出的拓扑(上面我们已说了这样的 $\tau$ 确可证明为一拓扑)是使 $f$ 为连续最弱(粗)的拓扑. 总之，若 $X$ 中已有拓扑 $\tau$，以 $f(U)$，$U\in\tau$，作为 $Y$ 中的开集，可以得到使 $f$ 为连续的最强的拓扑；若 $Y$ 中已有拓扑 $\sigma$，以 $f^{-1}(V)$，$V\in\sigma$ 作为 $X$ 中的开集，可以得到使 $f$ 为连续的最弱的拓扑.

现在得到了一个最重要的概念.

**定义 10** 设 $f:(X,\tau)\to(Y,\sigma)$ 是 (i) 一个一一映射，即一对一地把 $X$ 映到整个 $Y$ 上；(ii) $f$

是连续的；(iii) $f^{-1}$ 也是连续的，则称 $(X,\tau)$ 与 $(Y,\sigma)$ 同胚 (homeomorph)，$f$ 为一同胚映射 (homeomorphism).

注 我们用了一一映射 (bijection) 一词，常用的还有单射 (injection) 与满射 (surjection) 两个词. 前者指 $Y$ 中任一元 $y$ 如果有原像，则原像为唯一的；后者指 $Y$ 中一切点都是 $X$ 中某点的像，就是 $f$ 把 $X$ 映到 $Y$ 上，而不只是 $Y$ 内. 前者是说方程 $f(x)=y$ **如果有解，则解必是唯一的**；后者则说，方程 $f(x)=y$ 对 $Y$ 中的一切 $y$ **都有解**(但**不一定是唯一的**). 一个映射如果既是单射又是满射就称为是一一映射. 这时 $f(x)=y$ 对任意 $y\in Y$ 都有唯一解存在.

同胚的拓扑空间可以看成相同的空间，因为拓扑空间的基本定义就是开集，$X$ 与 $Y$ 同胚表明，不但 $X$ 与 $Y$ 作为点的集合，是通过 $f$ 实现一一对应的，而且 $\tau$ 与 $\sigma$ 作为子集族，也通过同一个 $f$ 实现一一对应. 如果粗略地说，$X$ 与 $Y$ 的一切拓扑性质都是以开集为基础来定义的，则在经过同胚映射后，开集是一一对应的，因而 $X$ 中开集的性质也都移到 $Y$ 中的开集上而没有改变. 把这一点说明白一点：所谓拓扑性质就是经同胚映射不变的性质. 因此，同胚的拓扑空间可以看成是相同的拓扑空间. 这一点经过下面的例子就看得很清楚，下面的几个例子在整个数学中起重要的作用，虽然限于 2 维情况，其重要性并不稍减.

**2. 拓扑空间的几个例子**

考虑 $n$ 维单位球面 $S^n$：

$$x_1^2+\cdots+x_{n+1}^2=1.$$

我们只考虑 2 维球面 $S^2$，并记上半球面 (不带边的) 为

$$S_+^2=\{x\in\mathbf{R}^3,x_1^2+x_2^2+x_3^2=1,x_3>0\}$$

(图 6-3-1) 它的闭包 $\overline{S}_+^2$，则是 $S^2\cap\{x_3\geqslant 0\}$. $S_+^2$ 的边界是平面 $x_3=0$ 上的单位圆周，即 $S^1$. 我们又记 $D^2$ 为平面 $x_3=0$ 上的单位圆盘：$x_1^2+x_2^2<1$，其边界即 $S^1$，而闭包 $\bar{D}^2$ 是闭圆盘 $x_1^2+x_2^2\leqslant 1$. 同样可以考虑下半球面 $S_-^2$. 在球面 $S^2$ 和平面 $\mathbf{R}^2$：$x_3=0$ 上都按我们直觉习惯的方式规定哪一些集合为开集，于是我们得到两个拓扑空间 $S^2$ 与 $\mathbf{R}^2$. 我们要看一下怎样稍作修改即可建立一种同胚关系.

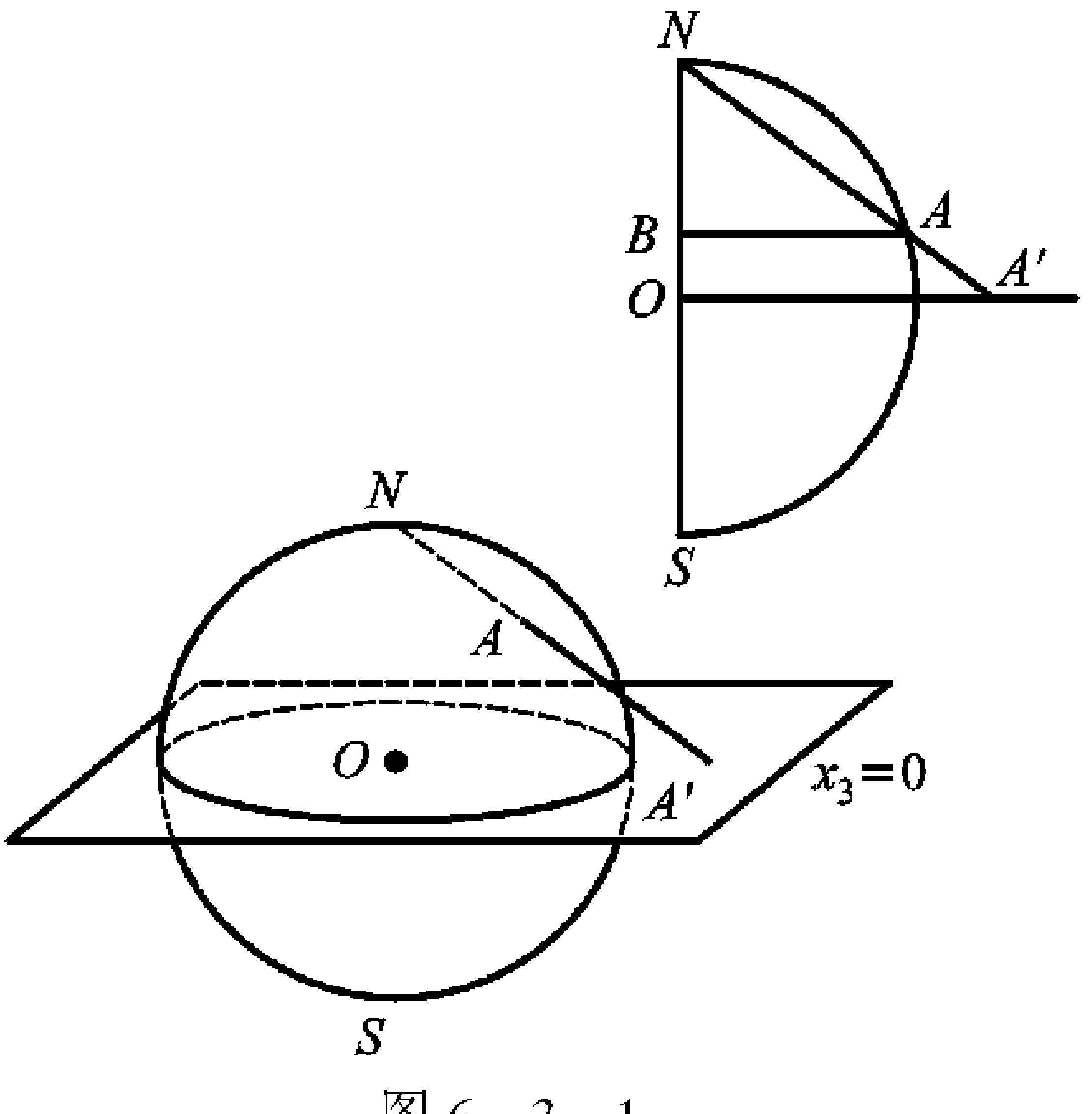


图 6-3-1

建立这样的同胚有多种方法，有一个最常见的称为球极射影 (stereographic projection)，其方法是由球面 $S^2$ 的北极 $N(0,0,1)$ 与球面 $S^2$ 上一点 $A$ 联直线，它交 $\mathbf{R}^2$ 于 $A'$ 点. 这里我们取 $\mathbf{R}^2$ 为赤道平面 $x_3=0$. 于是我们让 $A\sim A'$ 点. 直觉告诉我们，这个对应是一对一的，而且是双方 (即指映射 $A\to A'$ 与逆映射 $A'\to A$) 都是连续的，现在我们来计算一下. 以下凡是上加一撇“′”的都代表 $\mathbf{R}^3$ 中的平面 $\mathbf{R}^2$：$x_3'=0$ 上之对象，否则代表球面 $S^2$：$x_1^2+x_2^2+x_3^2=1$ 上之对象，相同的字母代表相对应的对象. 我们把 $NS$ 与 $NA$ 所决定的曲面与平面画在图 6-3-1 中，并设 $O$ 即原点，$ON$ 为 $x_3$ 轴，$OA'$ 为 $x_j$ 轴. 这里 $j=1,2$，而我们得到的结果也对 $j=1,2$ 均成立. 于是我们有 $ON=1$，$OA'=x_j'$，$BN=1-x_3$，$BA=x_j$，由于 $\triangle NOA'\backsim\triangle NBA$，所以

$$x_j' = x_j / (1 - x_3), j = 1, 2. \tag{2}$$

现求它的逆映射，将(2)平方后对 $j$ 求和有：

$$\begin{aligned}\| x' \|^2 &= \sum_{j=1}^{2} x_j'^2 = \sum_{j=1}^{2} x_j^2 / (1 - x_3)^2 \\ &= (1 - x_3^2) / (1 - x_3)^2 = \frac{1 + x_3}{1 - x_3}.\end{aligned}$$

这里我们利用了 $A \in S^2$，所以 $\sum_{j=1}^{2} x_j^2 + x_3^2 = 1$. 由上式得

$$1 - x_3 = 2 / (1 + \| x' \|^2). \tag{3}$$

由此即知 $A \sim A'$把 $\mathbf{R}^2$ 与除去 $N$ 点(即 $x_3 = 1$ 处)的球面 $S^2$ 双方一对一且双方连续地对应起来.

注意 $x_3 \to 1$ 的情况. 这时 $\| x' \| \to +\infty$. 如果我们认为 $\mathbf{R}^2$ 上当 $\| x' \| \to +\infty$时的点都是一个点，而且认为是无穷远点，或者说把这个“极限”看成一个点添加到 $\mathbf{R}^2$ 上，就得到“扩充的”$\mathbf{R}^2$：$\mathbf{R}^2 \cup \{\infty\}$. 而且我们就说扩充的 $\mathbf{R}^2$ 上只有一个无穷远点. 于是“扩充的”$\mathbf{R}^2$ 与 $S^2$ 是互相同胚的，球极射影(2)与(3)就是一个同胚映射. 于是按上面所讲的，$\mathbf{R}^2 \cup \{\infty\}$与 $S^2$ 应看成相同的拓扑空间. 例如 $S^2$ 上北极 $N$ 的邻域就变成扩充的 $\mathbf{R}^2$ 上无穷远点的邻域. 例如，$\| x' \| > R > 0$($R$ 是任意正数)就是无穷远点的一个邻域.

这个观点在研究复变量的函数时特别有用. 我们在第三、四章中讲过解析函数的微分与积分仍是解析函数，并且认为解析函数都定义在复平面 $\mathbf{C}$ 上的某个区域(甚至就是 $\mathbf{C}$ 本身)中，$\mathbf{C}$ 就是 $\mathbf{R}^2$. 所以我们是在未扩充的 $\mathbf{R}^2$ 上讨论解析函数的. 如果有某个函数 $f(z)$，令 $z = 1/t$，而得到函数 $f(z) = f(1/t) = \varphi(t)$，它在 $0 < |t| < \frac{1}{R}$中解析，如果 $\varphi(t)$一直到 $t = 0$ 都是解析的，我们就说 $f(z)$在 $z = \infty$处也是解析的. 这种观点可以说是在未扩充的 $\mathbf{R}^2$ 上把解析函数 $f(z)$解析延拓到 $\mathbf{R}^2 \cup \{\infty\}$上. 直到现在我们都是在未扩充的 $\mathbf{R}^n$ 讲微积分的. 现在许可采用另一种观点，即在扩充了的 $\mathbf{R}^2$，即 $\mathbf{R}^2 \cup \{\infty\}$上补充这个函数，而把无穷远真正看作一个点. 但这个观点与第一章绪论中讲的“实无限”还有区别. 如果看到，$\mathbf{R}^2 \cup \{\infty\}$其实就是 $S^2$，而北极 $N \in S^2$ 是其上一个“普通的”点，这是任何人都能接受的. 所以把$\{\infty\}$看成扩充的 $\mathbf{R}^2$ 上一个“普通的”点也是十分自然的事. 我们说这与“实无限”还有区别是因为讨论“实无限”时，我们甚至不需要找一个北极之类的对象来与之相应. 但是这里究竟会遇到 $t \to 0$ 时 $z \to \infty$，这个$\infty$究竟不是通常的数. 解决这个问题的办法是：不要奢求一下子就在整个 $S^2$ 上讨论函数而把 $S^2$ 分成几个部分：一部分不包括北极 $N$，例如就是 $\mathbf{R}^2$ 中的$|z| < R$，在这一部分里讨论函数时，我们取 $z$ 为自变量，或者说以 $z$ 为“局部坐标”. 另一部分则包括北极 $N$，也就是包括“扩充的”$\mathbf{R}^2$ 的“无穷远”点，例如$|z| > R_1$($R_1$ 与 $R$ 不一定相同而可取 $R_1 < R$)，这时我们就换一个新的局部坐标 $t$，而在$|t| < \frac{1}{R_1}$(注意，包括 $t = 0$，即 $z = \infty$)中讨论函数 $f(1/t) = \varphi(t)$. 这两个部分可能相重，例如在 $R_1 < |z| < R$ 中，我们则允许采用任一个局部坐标，而两种局部坐标 $t$ 与 $z$ 之间有一个“转换关系”(transition relation) $z = 1/t$. 这个十分自然的观点在现代数学中起了极大的作用，而拓扑的观点是其关键. 下一章我们将要详细地展开它. 至于在复变量函数的研究中，也就是在复分析中，在扩充的复平

面上讨论解析函数也是极富成果的.这个扩充的复平面就称为黎曼球.这是黎曼对整个数学的伟大贡献之一.

可能使读者感到意外的是,$\mathbf{R}^2$ 不仅有这一种扩充的方法. 为了解释这一点,我们从一个表面上看来全无关系,甚至显得有些不自然的问题开始.不过我们声明一点,由于避免篇幅过长,以及避免还要更多的准备知识,下面关于射影平面的讨论将更多地依靠直观,而不甚严格.但是这里讲的结果都是正确的,都有严格的解释和证明.

考虑 $\mathbf{R}^3$ 中过原点之所有直线之集合 $X$,我们把每一条这样的直线看成一个点,并用下面的方法对这些点的集合赋以拓扑,并讨论这样得到的拓扑空间.设一条这样的直线有方向数 $(\alpha_1, \alpha_2, \alpha_3)$,这些 $\alpha_i$ 不能同时为 0:$\sum\limits_{i=1}^{3} \alpha_i^2 \neq 0$,于是,任取一个非零实数 $\lambda\,(\neq 0)$,$(\lambda\alpha_1, \lambda\alpha_2, \lambda\alpha_3)$ 仍是同一条直线的方向数,所以我们不能说每一条这样的直线 $l$ 必与唯一一组 $(\alpha_1, \alpha_2, \alpha_3)$(其中 $\sum\limits_{i=1}^{3} \alpha_i^2 \neq 0$)一一对应,而应该在这些三元组之间建立一个等价关系:

$$(\alpha_1, \alpha_2, \alpha_3) \sim (\beta_1, \beta_2, \beta_3) \Leftrightarrow \exists\, \lambda \neq 0 \text{ 使 } (\alpha_1, \alpha_2, \alpha_3) = \lambda\,(\beta_1, \beta_2, \beta_3),$$

这里 $\alpha = (\alpha_1, \alpha_2, \alpha_3)$,$\beta = (\beta_1, \beta_2, \beta_3)$ 都属于 $\mathbf{R}^3 \setminus \{0\}$,这些 $\alpha, \beta$ 称为这条直线的齐次坐标(homogeneous coordinates)——其实并不是严格意义下的坐标,因为直线 $l$ 与方向数 $\alpha$ 并非一一对应,它只与 $\alpha$ 在关系 $\sim$ 下的等价类一一对应.这些等价类之集合我们记作 $\mathbf{R}^3/\sim$,称为 $\mathbf{R}^3$ 对于 $\sim$ 的商空间(quotient space). 商空间的概念是数学的基本概念之一. 其实我们过去已见过很多. 例如周期运动,一个质点沿单位圆周作旋转,其所走的路程当然构成整个实数域 $\mathbf{R}$,但是每转一圈(正向或反向),这个质点都回到原来的位置,所以如果不问它已转了多少圈,而只问它现在在圆周上的位置,则运动路程 $S$ 和 $S + 2k\pi$, $k \in \mathbf{Z}$ 其实代表同一位置——或称相位(phase).所以我们用 $S \sim S + 2k\pi$ 表示这个情况. 这个质点走过的路程 $S \in \mathbf{R}$,但其相位之集合则为 $\mathbf{R}/\sim = \mathbf{R}/2\pi\mathbf{Z}$.相位空间是路程空间关于旋转整数周的商空间.现在回到原来的问题.在 $X$ 的每个元,即一条过 $O$ 的直线 $l$ 的等价类中选一个代表元.为此,在 $\mathbf{R}^3$ 中以 $O$ 为心作单位球面 $S^2$: $x_1^2 + x_2{}^2 + x_2{}^3 = 1$,于是每一个 $l$ 均与它交于两点,设其一是 $x$(用交点在 $\mathbf{R}^3$ 中的直角坐标来表示),则另一交点是 $-x$,这两点位于一条直径的两端,称为对径点(antipodal points).我们可以有两个说法,一是在这两个对径点中任取一点为其代表元,例如取位于下半球面 $S_-^2$ 上的一点代表 $l$,但这时位于赤道平面上的一对对径点,例如图 6-3-2 中的 $a$ 与 $a'$ 又如何?我们也可以规定取前面的一点.或者我们把 $S^2$ 当成地球表面,认为 $NA$ 为始初子午线,并由它开始计算经度,并且取经度在 $[0, \pi)$ 中的点为代表元,这样 $X$ 中的每一条直线均与 $S_-^2 \cup$ {前半条赤道}上的点一一对应,我们记 $S_-^2 \cup$ {前半条赤道}为 $\widetilde{S}_-^2$,则 $X$ 与 $\widetilde{S}_-^2$ 作为集合是一一对应的.这样的说法固然直观,但是太多人为而不自然的痕迹.尤其是,在应用它作具体计算时很不方便,因此我们又想到利用商空间的技巧,即认为 $x$ 与 $-x$ 互相等价:$x \sim (-x)$,而 $X$ 与 $S^2/\{x \sim (-x)\}$ 一一对应.对此有一个重要的说明:第一种说法的好处仍然在于其直观性,回到上面讲的绕圆旋转的例子,起点位于 $S = 0$ 与 $S = 2\pi$,其实是一个点,就是说,如果把线段 $[0, 2\pi]$ 的两端"粘合"起来,它就是圆周. $S = 0$ 与 $S = 2\pi$ 是等价的:具有相同相位,我们就把它们"粘合"在一起. 把这个想法用于我们的例子:首先把 $S_+^2$ 去掉,只看下半球面 $S_-^2$,然后再把下半球面的边缘——赤道——上所有的

对径点,例如图 6-3-2 中的 $a$ 与 $a'$,粘在一起,这样就可以得到 $\widetilde{S}^2_-$,而仍有 $X$ 与 $\widetilde{S}^2_-$ 作为集合一一对应.到这里读者会问,具体怎样粘法?举一个日常生活中的例子,如果你擀出了一张圆形的饺子皮,为了准备包馅,先把它拢成碗形,我们说这就是 $S^2_-$,可是最后怎样按上述要点捏成一个饺子呢?——要求饺子皮边上的每一对对径点要捏在一起.我们直觉的经验是:这是捏不成的.您可以做一个黎曼球那样的"饺子",但做不成 $S^2_-$ 饺子. 为什么做不成?数学上可以证明,只有在 4 维空间才可以做成. 这里我们回避了许多重要的数学问题:什么是商空间?它有什么性质?什么叫在 4 维空间才能做成?这些属于拓扑学的范围. 但是为了看出它的困难所在,我们不妨再退一步,如果只是部分地把对径点捏到一起了,例如已把 $S^2_-$ 边缘上用虚线画的两段粘在一起了:$a$ 点粘住了 $a'$,$b$ 点粘在 $b'$ 上,这成为一个什么?把"饺子皮"$S^2_-$ 拉成一条长方形,把它的右端 $a'b'$ 转 180°后粘到左边的 $ab$ 上就得到下面那个著名的曲面——默比乌斯带(Möbius strip)(图 6-3-3).如果上面我们不把 $a'b'$ 转 180°,而允许 $a'$ 粘到 $b$ 点,$b'$ 粘到 $a$ 点,就会得到一个柱面,它有上、下两个边,是两条曲线,可是默比乌斯带的边只是一条曲线.请问,怎样把余下的部分 $ab'$ 与 $a'b$ 按我们的要求粘在一起呢?

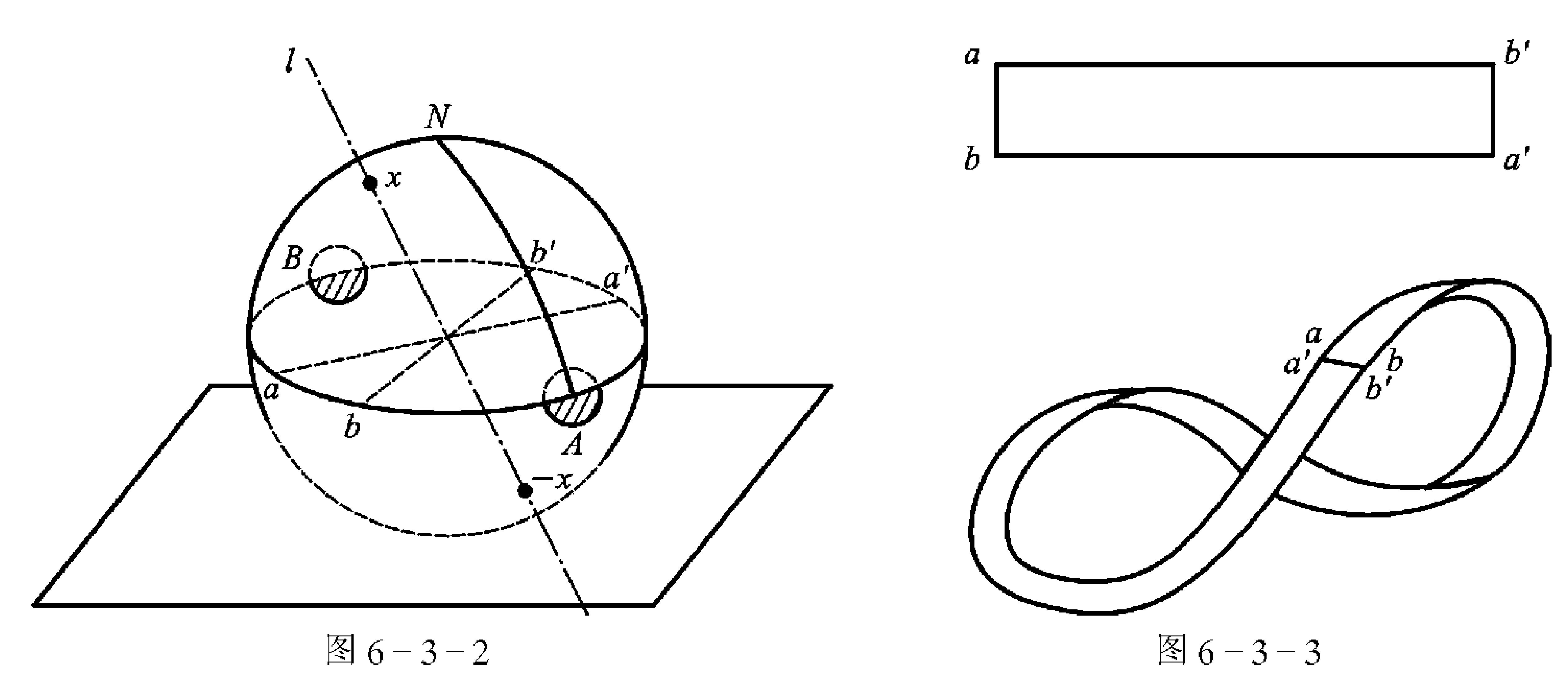



图 6-3-2　　图 6-3-3

回忆一下,我们在第二章中介绍了黎曼曲面,那也是一种无法在我们朝夕生活于其中的 $\mathbf{R}^3$ 中实现的. 数学中出现了这样一些研究对象,并不只是由于想把微积分的基本概念弄得更明白,例如黎曼提出黎曼曲面是与他研究所谓阿贝尔积分密切相关的. 更进一步,他终生思考的问题是,我们生活于其中的空间的物理本性究竟是什么?对于空间的研究与物理学的许多重大问题相联系.拓扑学的创始者之一庞加莱研究拓扑学开始是出于天体力学的所谓 $n$ 体问题,于是非常自然地出现了种种奇特的空间结构.正是为了研究这些空间结构,才进一步感到必须把一些最基本的概念弄清楚.这已经是 20 世纪初年的事了.

现在再回到上述关于 $X$ 与 $S^2_-$ 的研究.我们可以再提出一个与它们同胚的空间.在 $S^2$ 的南极 $N$ 处作一个切平面 $x_3=-1$,在这个平面与 $\widetilde{S}^2_-$ 间先建立一个作为集合的一一对应:若在 $S^2_-$ 上有一点 $P$,联结 $OP$ 交 $x_3=-1$ 于 $P_1$(见图 6-3-4),反之,若在 $x_3=-1$ 上有一点 $P_1$,联结 $OP_1$ 交 $S^2_-$ 于 $P$.这样我们令 $P$ 与 $P_1$ 对应即得 $S^2_-$ 与 $x_3=-1$ 的对应.注意,这个对应还没有拓展到 $\widetilde{S}^2_-$ 上,因为联结 $O$ 与赤道上一点例如 $a$ 点,将得到一条与 $x_3=-1$ 平行的直线,它与 $x_3=$

$-1$ 不会在任意有限远处相交.所以,若想把这个一一对应拓展到整个 $\widetilde{S}_-^2$ 上,就需要在此平面上添加一些理想元素,就如同把复平面变成扩充复平面要增加一个无穷远点 $z=\infty$,使它成为黎曼球一样.但是我们不能简单地在 $x_3=-1$ 上再添加一个点.因为在扩充复平面时,我们的目的是要把复变量的解析函数推广到 $z=\infty$ 附近,为此我们要利用 $z=1/t$ 以保持 $f(z)$ 的解析性质,而在变换 $z=1/t$ 之下,扩充的复平面上只许可有一个无穷远点 $z=\infty$ 与 $t=0$ 相应.现在我们拓展 $x_3=-1$ 的目的是什么呢?我们希望拓展后的平面尽可能多地保持欧氏平面的性质.在欧氏平面上有两个互相对偶的性质:(i)经过两个不同点可以作一条且仅有一条直线联线.(ii)两条直线,如果不平行,则必有一个且仅有一个有限远处的交点.请注意,第二个提法就是著名的平行线公设.这个公设在欧几里得《几何原本》中表述如下:"若一直线落在两直线上所构成的同旁内角和小于两直角,那么把两直线无限延长,它们将在同旁内角和小于两直角的一侧相交."直到1795年,普赖菲尔(John Playfair)才把这个公设改述为我们熟悉的形式:过直线外一点可以作一条,且仅可作一条与此直线平行的直线.所谓平行直线,《几何原本》卷Ⅰ之定义23是说,即"向两方无限延伸,而在两个方向上彼此不相交的直线."可见,欧几里得是非常谨慎地避免说到"在无穷远处相交"一类的话,因为什么是"无穷远处"在数学上是没有定义的.现在我们扩充平面 $x_3=-1$ 想达到的目的就是把上述(i)、(ii)两条加以推广,使(i)成为:(i′)经过两个不同点(不论是有限远点还是无穷远点)都可以作一条且仅有一条直线.(ii)则推广成:(ii′)两条不同直线必有一个且仅有一个交点(如果它们不平行,这个交点就是有限远点,如果平行,这个交点就是无穷远点).这样一来,"平行线必在无穷远处相交"这个原本含义模糊的命题就变得完全确切清晰了.一个平面,这样添加上无穷远点以后,称为射影平面,记作 $RP^2$.射影平面上的几何学称为射影几何学.我们立刻就会看到,射影平面与黎曼球面是截然不同的数学对象.由此可见,为了达到我们的目的,首先要说明什么是 $x_3=-1$ 上的直线.如果在 $x_3=-1$ 上已经有了一条我们熟知的有限远处的直线(或线段)$l$(或 $P_1Q_1$),则 $OP_1$,$OQ_1$ 决定 $\mathbf{R}^3$ 中的一个平面,它与半球面 $S_-^2$ 沿一大圆弧 $aPQb$ 相交.大圆弧(劣弧)$PQ$ 对应于 $P_1Q_1$.在 $x_3=-1$ 上我们遇到了什么是无穷远点的困难,可是在球面 $S^2$ 以及 $S_-^2$ 上却没有这个困难.因此,我们把直观而朴素的直线"定义"推广为:$x_3=-1$ 上,大圆劣弧的像就是直线,大圆由通过球心的平面决定,大圆就是这个平面与球面之交.图 6-3-4 上我们画了一条通过有限远点 $P_1$ 的直线 $l$,它决定了相应大圆的平面,即 $OPQ$ 平面,它与赤道平面交于直径 $ab$.如上所说,$a\sim b$,而且我们以 $b$ 为代表元($b\in\widetilde{S}_-^2$),$b$ 既然在大圆上,$b$ 之像就应该在 $l$ 上.在相应于 $l$ 的大圆劣弧 $(a,b]$($a$ 不算弧上的点,但是 $b$ 算)上,只有这一点在赤道平面上,它的像不在有限远处.因此,我们定义 $b$ 的像即 $l$ 上的无穷远点.这与我们的直观是很符合的:直线 $Ob$ 与平面 $x_3=-1$ 平行,所以它与此平面交于无穷远点.但是从图 6-3-4上看,由 $P_1$ 沿 $l$ 向 $Q_1$ 移动时,球面上的相应点向 $b$ 移动,因此趋向无穷远.但由 $Q_1$ 向 $P_1$ 移动时,球面上相应点也趋向赤道平面,所以在 $l$ 上也趋向无穷远点,那么是否在 $l$ 上有两个无穷远点呢?没有.因为后一情况下球面上的点趋向 $a$,而上面已经说过,$a$ 与 $b$ 是粘合在一起的:$a\sim b$,所以不论是向 $Q_1$ 移动还是向 $P_1$ 移动,都会趋向 $l$ 上的同一个无穷远点.总之我们看到,在 $x_3=-1$,即在 $RP^2$ 上,每条通过至少一个有限远点 $P$ 的直线 $l$ 上均有恰好一个无穷远点,即大圆与赤道交点 $b$ 的像.上面我们加上了一点限制:"通过至少一个有限远点 $P$ 的直线 $l$".

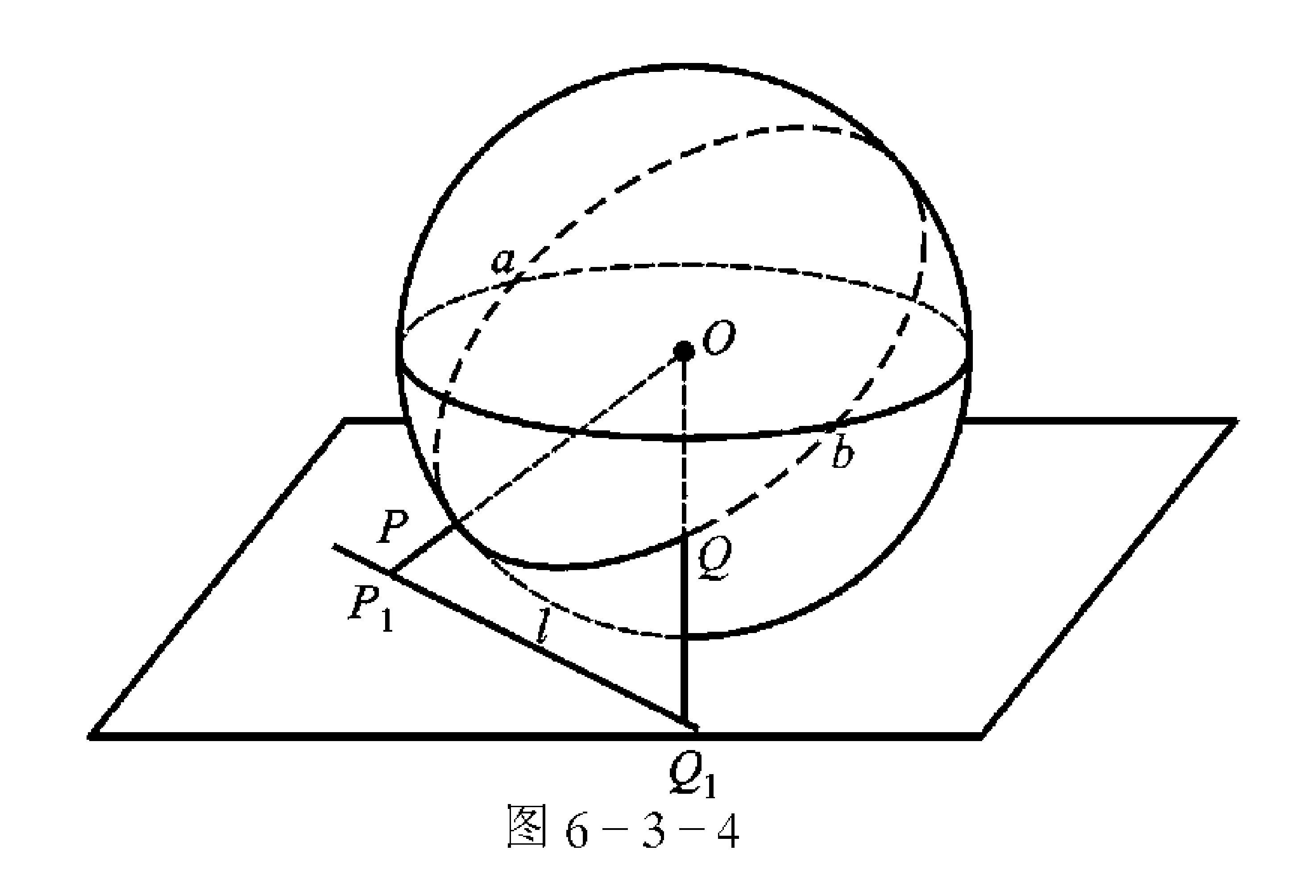


图 6-3-4

那么有没有不通过任何有限远点的直线呢？注意到，每一条直线都相当于一个大圆（其上的对径点视为一点），则还有一个大圆的像不在有限平面上，这就是赤道平面上的大圆（或如我们前面说的，只是这个大圆上经度在 $[0,\pi)$ 上的一段）. 所以我们在平面上再引入一条理想的直线，称为无穷远直线. 因为其上的点都是某一直线 $l$ 上的无穷远点，而没有有限远点. 这样一来，通过两个不同无穷远点也有一条直线，即赤道平面上的大圆，这条直线就是唯一的不通过任何有限远点的直线. 总之我们看到，为了保证 $\widetilde{S}^2_-$ 与平面 $x_3=-1$ 的点一一对应，就应该在平面上增添一些"理想元素"：每条有限直线（即通过至少一个有限远点的直线）上加上一个（且仅加一个）无穷远点以对应于赤道平面上之大圆上的点；整个平面应加上一条无穷远直线（它不通过任何有限远点），以对应于赤道平面上之大圆. 添加了这些理想元素的平面称为射影平面 $RP^2$.

我们还可以多说几句：每条有限直线添一个无穷远点，这在直观上好理解，可是这些无穷远点的集合何以要称为一条直线，而不简单地就说是"无穷远点轨迹"之类？一方面因为这个轨迹是赤道平面上之大圆的像，而大圆之像总应该说是直线. 还有另一个理由：我们一开始就讨论 $\mathbf{R}^3$ 中通过原点的直线，并且对每一条这样的直线均给以"齐次坐标"，即其方向数，而方向数 $(\alpha_1,\alpha_2,\alpha_3)$ 之三个分量不许同时为 0. 其实在射影几何中，平面上每个点 $(x,y)$ 也都有一个齐次坐标：令 $x=x_1/x_3$，$y=x_2/x_3$，$(x_1,x_2,x_3)$ 就称为此点的齐次坐标，我们也规定这三个分量不得同时为 0. 我们知道，平面上的直线方程是

$$a_1x+a_2y+a_3=0,$$

其中 $a_1,a_2$ 不许同时为 0，但引入齐次坐标后，它成为

$$\sum_{i=1}^{3} a_ix_i=a_1x_1+a_2x_2+a_3x_3=0. \tag{4}$$

因为 $(x_1,x_2,x_3)$ 地位平等——即地位对称——我们不应对哪一个有所偏爱，所以这时应加的限制就只能是 $a_1,a_2,a_3$，不得同时为 0（否则 (4) 式成为平凡的恒等式）. 若 $a_1=a_2=0$，$a_3\neq 0$，则 (4) 成为

$$x_3=0, \tag{5}$$

它仍是一条直线（一次方程），称为无穷远直线. 同样，一个点 $(x,y)$ 即 $(x_1,x_2,x_3)$ 中，本应 $x_3\neq 0$，但由于这三个分量现在是对称的，我们不应"歧视" $x_3$，不应该不许它为 0，而应规定 $x_1,x_2,x_3$ 不能同时为 0（对方向数也有类似规定，若 $\alpha_1=\alpha_2=\alpha_3=0$，就好比只有一个原点，作不出任何直

线了). 我们规定,凡齐次坐标为 $(x_1, x_2, 0)$ 之点称为无穷远点,当然它们的集合是无穷远直线(5). 从这一段讨论可以看到,无穷远点、无穷远直线与有限点、有限直线比较,没有任何本质上的特殊之处. 这正是射影几何学的精义之一.

有了这些讨论,就可以看到欧氏平面上那两个对偶的性质现在成为

(i′) 过任意两个不同点必有唯一直线.

当这两点都是有限远点(如图 6－3－4 上的 $P_1Q_1$),这直线就是 $S^2_-$ 上的大圆弧 $PQ$;若一点是有限远点 $P$,一是无穷远点 $b$,就是过 $P$ 与 $b$ 的大圆弧,它的像即过 $P_1$ 而与直径 $a$, $b$ 平行的直线;若两点都是无穷远点,这直线就是无穷远直线.

(ii′) 任意两条不同直线必有唯一交点.

这是因为两个大圆必有交点,也是因为两个不同的方程(4) $a_i x_1 + b_i x_2 + c_i x_3 = 0$, $a_i^2 + b_i^2 + c_i^2 \neq 0$, $i = 1, 2$, 且 $(a_1, b_1, c_1)$ 与 $(a_2, b_2, c_2)$ 不等价,必有非 0 的公共解. 如果这个交点不在赤道平面上而是有限远点,这就是初等几何中讲的两直线相交于某有限远点. 如果这个交点是赤道平面上某点例如 $b$ 点,则从初等几何看来,这两条直线必平行于直径 $ab$,而有相同的无穷远点 $b$ 为其公共点. 这一点读者很容易用立体几何的简单知识证明.

那么普赖菲尔的平行公设怎么说呢? 如果平行线理解为不相交的直线,那么这个公设错了,因为它与(ii′)矛盾. 如果把该公设理解为从 $l$ 外一点必有一直线与 $l$ 相交于 $l$ 上的无穷远点 $b$,那么它就是(i′). 总之,在射影几何中没有给"平行性"留下特殊的地位,或者干脆就说平行性不是射影几何的概念.

倒是有一点应该强调,(i)与(ii)是对偶的:把"点"与"直线"互换一下,(i)就变成(ii),(ii)就变成(i). 射影几何中许多最基本的定理都是对偶的,只要证明了其一,把其中的点与直线互换就得到另一个对偶的定理. 初等几何中的许多定理都有这样的性质. 而在现代数学中点 $x$ 与方向数所代表的向量 $\xi$ 的对偶关系,起了十分重要的作用. 我们在第五章中还联系着量子力学讨论了这种对偶关系. 这种对偶关系和射影几何中的对偶关系本质上是相通的.

我们似乎离题太远,这是有原因的. 射影几何起源于透视,这在建筑学与绘画中都是基本的. 在 19 世纪中,射影几何成了数学发展的主流的一个突出的部分. 当然,时过境迁,现在可以说是风光不再了. 然而,它的许多基本概念都成了数学中不可少的财富. 由于它与默比乌斯带等等有密切关系,而默比乌斯带在许多问题上又是典型,如它的单侧性、不可定向性(在下一章我们将看到这些概念在微积分中是多么重要),所以,对射影几何的一些基本概念还是知道一些为好. 因为现在大学的数学课程一般都不再讲有关内容了,我们借此机会在这里多占用一些篇幅讲了以上的内容. 下面还是回到拓扑学的问题.

我们上面实际上是讲了三个集合即 $X$($\mathbf{R}^3$ 中过原点之直线的集合),$\widetilde{S}^2_-$ 以及 $RP^2$ 之间有一一对应关系. 但是重要的是如何在其上建立拓扑,从而使得这些集合之间的一一对应变为拓扑空间的同胚. 这里最容易的从 $S^2$ 开始,$S^2$ 上如何定义开集似乎直观上没有什么困难. 我们也可以一言以蔽之:让 $S^2$ 赋有 $\mathbf{R}^3$ 的子空间拓扑,即 $U \subset S^2$ 为一开集,当且仅当有一个 $\mathbf{R}^3$ 的开集 $V$,使 $U = V \cap S^2$. 进一步我们来定义 $\widetilde{S}^2_-$ 上的拓扑,由于 $S^2_-$ 是 $S^2$ 在一种等价关系 $\sim$ 下的商空间,我们说 $U \subset \widetilde{S}^2_-$ 是一个开集的意义如下:任找一个 $x \in U$,它在 $S^2$ 中所定义的等价类是 $S^2$ 的一

个子集$\{x, -x\}$，把这些子集(注意，不是看作一个等价类而是看成一个子集)并起来，得到一个$V$(用一个很容易看懂的写法：$V = U\cup(-U)$)，如果$V$是$S^2$的一个开集，我们就说$U$是$\widetilde{S}^2_-$的一个开集，这就是拓扑学中在商空间中定义商拓扑的方法. 这件事看来简单，但有一些细微之处要注意，例如看图6-3-2，在$A$点附近的$U_-$，它在$S^2$中的等价类之并分成两块，一块是$S^2_-$上$A$附近的$U_-$，另一块是"下半球面"$S^2_-$上$B$点附近的$U_+$(图上的阴影区域). 如果$A$点附近的$U_-$是包括了赤道大圆弧的话，则$B$附近的$U_+$不应包括赤道大圆弧. 这两块并起来并非$S^2$中的开集. 但若在$\widetilde{S}^2_-$中取$U_-\cup U_+$而且包括赤道大圆弧，即图6-3-2中分离的两块$U_\pm$，它们一在$A$点附近，一在$B$点附近，则它在$S^2$中的等价类之并合成两块，即图6-3-2上的两块连通的$U_-\cup U_+$. 它们都跨过赤道大圆，而且都是$S^2$上的开集，因此其并也是开的，所以$\widetilde{S}^2_-$上分离的两块$U_\pm$合在一起算是$\widetilde{S}^2_-$的开集. 有了$\widetilde{S}^2_-$上的拓扑，即可定义$RP^2$上与$X$上的拓扑. 我们来看前者，如果记图6-3-3上由$O$到$P$到$P_1$之映射为$\varphi$，我们说$U\subset RP^2$为一开集，如果它是$\widetilde{S}^2_-$上某开集$V$的像的话：$U=\varphi(V)$. 按前面的说法，这样在$RP^2$上定义拓扑乃是保证$\varphi$为连续映射的最强的拓扑，但同时也是使$\varphi^{-1}: RP^2\to S^2_-$为连续的最弱的拓扑. 所以这是使$\varphi$为同胚(即$\varphi$与$\varphi^{-1}$均为连续)的唯一的拓扑. 对$X$也可同样规定其拓扑.

总之，我们得到的不只是$X$，$\widetilde{S}^2_-$与$RP^2$作为集合一一对应，而且是它们作为拓扑空间的同胚. 这三个空间实际是相同的，是射影平面的三个不同的表示方法.

把它与黎曼球比较就会感到，原来两个局部地看来完全一样的拓扑空间(二者局部地看都是$\mathbf{R}^2$的一部分)，整体地看却完全不同. 这样不由得提出一个问题：给出两个拓扑空间$X$与$Y$怎样看出它们是否同胚的呢？这是拓扑学的根本问题，也是无法完全解决的问题.

讲完了关于拓扑空间及其映射的基本概念以后，读者可能会想到，是否应该继续讨论有关序列极限等等问题了. 很遗憾，不可以. 通常凡讲到序列问题时，总是联系着度量空间的. 关于序列的一些基本定理在一般拓扑空间中并不成立. 当然，在微积分中有一些极限是不能归结为序列极限的，例如在第四章中就讲到，黎曼积分的积分和的"极限"应该说成是上、下和的下、上确界更好. 为了讨论这一类极限问题，以及把极限与收敛性问题放在拓扑空间来讨论，就需要引入"网"(net)的概念. 这是拓扑学的另一篇章，我们在此割爱.

**3. 紧拓扑空间** 现在回到拓扑空间的两个重要的特性：紧性与连通性. 在这里我们将看到一般的拓扑空间与前面讲过特殊的拓扑空间——度量空间有多么大的差别. 同时也把一些可以在上一节就开始讨论而没有展开的问题在这里继续下去. 但在这之前，我们先介绍一个似乎不起眼而又十分重要的概念——可分离性.

**定义 11** 若$(X, \tau)$为一拓扑空间，若$X$中任意两个不同点$x_1$与$x_2$，均有其邻域$U_i$ $(i=1, 2)$，使$U_1\cap U_2=\varnothing$，这时$X$称为豪斯多夫空间(Hausdorff space)，或可分离空间.

这个概念似乎是完全平凡不足道的，它无非说$X$中的两个不同点$x_1$与$x_2$总是可分离的. 在度量空间中它必然成立(读者可以自己证明)，但在一般的拓扑空间中它确实可能不成立. 举一个平凡不足道的例子，若$\tau$是平凡的拓扑，即只有$\varnothing$与$X$两个开集，这种不相交的邻域就找不到了. 可是在数学中确实还有非平凡的拓扑空间也是非豪斯多夫的，这说明拓扑空间确实是一个

极广泛的概念，我们有时不得不稍加限制使它更接近我们熟悉的度量空间. 引入这个概念目的在此. 附带提一句，引入 $(X,\tau)$ 的可分离性还是因为下一章将讨论的微分流形必须是可分离的. 若不然，微积分中许多重要工具都会失效.

下面我们开始讨论紧性. 对于一般的度量空间，§1 中已经讲了的关于 $\mathbf{R}$ 的紧性的结果（§1，定理 5）现在要作为定义来用.

**定义 12** 设 $(X,\tau)$ 为一拓扑空间，若 $X$ 之任一开覆盖 $\{U_\lambda\}$ 中必有一个有限子覆盖，则称 $(X,\tau)$ 为紧空间. 若 $A\subset X$ 为一子集，且 $A$ 作为 $(X,\tau)$ 之子空间是紧的，则 $A$ 称为紧集.

**注** 上面说 $A$ 作为子空间即指 $A$ 上赋有由 $X$ 诱导而得的子空间拓扑. 所以当我们说 $\{U_\lambda\}$ 是 $A$ 的开覆盖时，$U_\lambda$ 就是子空间拓扑中的开集. 亦即存在 $(X,\tau)$ 中的开集 $V_\lambda$，使 $U_\lambda=V_\lambda\cap A$. 因此，在许多书中讲 $A$ 是紧集时说："若有 $A$ 的开覆盖 $\{V_\lambda\}$，其中 $V_\lambda$ 是 $X$ 中之开集，则必有 $A$ 的有限子覆盖"，这与我们的讲法当然是一致的.

这个定义当然地适用于度量空间.

**定理 6** 若 $f:(X,\tau)\to(Y,\sigma)$ 为连续，$A\subset X$ 为紧集，则 $f(A)\subset Y$ 也是紧集.

**证** 设 $\{V_\lambda\}$ 是 $f(A)$ 的开覆盖（这里的用法是指 $V_\lambda$ 是 $Y$ 中的开集），由于 $f$ 是连续的，故 $f^{-1}(V_\lambda)=U_\lambda$ 是 $(X,\tau)$ 的开集，而且 $\{U_\lambda\}$ 是 $A$ 的开覆盖，因为 $A$ 是紧集，故必有 $\{U_\lambda\}$ 的一个有限子集 $\{U_1,\cdots,U_N\}$ 使 $A\subset\bigcup_{k=1}^{N}U_k$，从而 $f(A)\subset\bigcup_{k=1}^{N}f(U_k)=\bigcup_{k=1}^{N}V_k$，即是说 $\{V_1,\cdots,V_N\}$ 是 $f(A)$ 的有限子覆盖，从而 $f(A)$ 为紧. 证毕.

**注** 如果 $A=X$，这个定理仍成立.

紧集在连续映射下之像仍为紧，但 $(Y,\sigma)$ 中的紧集 $K$ 在 $(X,\tau)$ 中的原像则不一定为紧. 如果对 $Y$ 之任一紧子集 $K$，$f^{-1}(K)\subset X$ 仍为紧，则 $f$ 称为适当的 (proper).

紧集在同胚映射下之像当然仍是紧的. 这个简单的事实却告诉我们一个重要事实：若 $X=Y=\mathbf{R}$，$A=[a,b]\subset X$，$B=(a,b)\in Y$，则 $[a,b]$ 与 $(a,b)$ 决不可能同胚，否则 $(a,b)$ 将成为 $\mathbf{R}$ 的紧子集. 但在 §1 中我们已经看到实数系的紧子集必为有界闭集，而 $(a,b)$ 不是有界闭集，所以它与 $[a,b]$ 不可能同胚. 由此看来，开区间与闭区间的拓扑性质有很大的区别. 到高维 $\mathbf{R}^n$ 时，情况也一样. 所以在整个微积分理论中，我们时时都要注意是在开集还是在闭集上讨论问题. 最简单如函数 $f(x)$ 的连续性，在闭区间上讨论时我们习惯把连续性了解为单侧连续性，只是最简单的例证.

紧性与闭性有密切的关系，我们有

**定理 7** 紧空间 $X$ 的闭子集 $A$ 必为紧集.

**证** 设 $A$ 有一个开覆盖 $\{U_\lambda\}$，因为 $A$ 为闭，所以 $A^c=X\setminus A$ 是 $X$ 的一个开子集，记作 $U^0$，于是 $\{U_\lambda,U^0\}$ 构成 $X$ 的开覆盖，而存在一个有限子覆盖. 不失一般性，我们可把 $U^0$ 也列入这个子覆盖中，而得有限子覆盖为 $\{U_1,\cdots,U_N,U^0\}$，所以 $X=U^0\cup\left[\bigcup_{k=1}^{N}U_k\right]$. 所以上式右方亦必覆盖 $A$. 但 $U^0\cap A=\varnothing$，所以 $A\subset\bigcup_{k=1}^{N}U_k$，即是说从 $\{U_\lambda\}$ 中可以找到 $A$ 的有限覆盖 $\{U_1,\cdots,U_N\}$，所以 $A$ 是紧集.

这个定理的逆则不太一样了.

**定理 8** 豪斯多夫空间的紧子集 $A$ 必为闭.

**证** 为证 $A$ 是闭集,只需证明任一点 $x\in A^c=X\setminus A$ 必有一个与 $A$ 不相交的邻域 $V$ 即可.为此任取一点 $z\in A$,由于 $X$ 是豪斯多夫空间,故必有 $x$ 与 $z$ 的开邻域 $V_x$ 与 $U_z$ 互不相交:$V_x\cap U_z=\varnothing$.但 $\{U_z\}_{z\in A}$ 是 $A$ 的一个开覆盖,故必有一个有限子覆盖 $\{U_1,\cdots,U_N\}$ 使 $A\subset\bigcup_{k=1}^{N}U_k=U$.相应地则有 $x$ 的有限多个开邻域 $\{V_1,\cdots,V_N\}$ 使 $V_i$ 与 $U_i$ 不相交.记 $V=\bigcap_{k=1}^{N}V_k$,易见 $x\in V$,$A\subset U$ 而且 $V\cap U=\varnothing$.事实上,若有某点 $y\in V\cap U$ 则 $y$ 必属于某个 $U_j$,同样 $y\in V=\bigcap_{k=1}^{N}V_k\subset V_j$,于是 $V_j\cap U_j$ 中有至少一个元 $y$ 而非空集,这与 $V_x\cap U_z=\varnothing$ 相矛盾.

上面我们实际上证明了比定理所宣布的更多的结果:不但 $x$ 有一个邻域与 $A$ 不相交,而且还与 $A$ 的某个邻域 $U$ 不相交,而这又是下面结果的特例.

**定理 9** 设 $X$ 是豪斯多夫空间,$A_1$,$A_2$ 为 $X$ 之互不相交的紧子集,则必存在 $A_1$,$A_2$ 之邻域 $W_1$ 与 $W_2$,使 $W_1\cap W_2=\varnothing$.

**证** 任取一点 $x\in A_1$,用定理 8 的证法,必可找到 $x$ 的邻域 $V_x$(即上面证明中的 $\bigcap_{k=1}^{N}V_k$)与 $A_2$ 的邻域 $U_x$(即上面的 $\bigcup_{k=1}^{N}U_k=U$),使得 $V_x\cap U_x=\varnothing$,而 $A_2\subset U_x$,但是 $\{V_x\}_{x\in A_1}$ 构成 $A_1$ 的开覆盖,所以可以由其中选出有限个 $V_x$ 来:$\{V_1,\cdots,V_M\}$,使得 $A_1\subset\bigcup_{k=1}^{M}V_k=W_1$,在 $\{U_x\}_{x\in A_2}$ 中取出相应的 $\{U_1,\cdots,U_M\}$,则 $A_2\subset\bigcap_{k=1}^{M}U_k=W_2$,$W_2$ 是 $A_2$ 的开邻域,而且仿照上面的证法,$W_1\cap W_2=\varnothing$.定理证毕.

这个定理告诉我们,一个空间若为豪斯多夫空间,则不但其不同点是互相分离的,而且其不相交的紧子集也是互相分离的.

至此,我们试图来刻画一个拓扑空间的全部紧子集.最简单的情况当然是 $\mathbf{R}^n$,我们可以证明,$\mathbf{R}^n$ 中的全部紧子集,即全部有界闭集.所以早期一点的微积分教材时常不讲紧性这个概念,而只讲其有界闭集如何如何.不言而喻,$\mathbf{R}^n$ 必是豪斯多夫空间.

**定理 10** $\mathbf{R}^n$ 的紧子集,即 $\mathbf{R}^n$ 的有界闭集.

**证** 我们只对 $n=1$ 来证明,先证 $\mathbf{R}^1$ 中的有界闭集 $A$ 为紧集.这在本质上就是 §1 定理 6(海涅-博雷尔定理),不过那里讲的只是闭区间 $[a,b]$ 而不是一般的有界闭集 $A$,所以要稍作修改.因为 $A$ 是有界的,所以必有一个有界闭区间 $[m,M]$,使 $A\subset[m,M]$.令 $A^c=[m,M]\setminus A$,则 $A^c$ 是子空间拓扑中的开集.设有 $A$ 的开覆盖 $\{U_\lambda\}$,则 $\{U_\lambda,A^c\}$ 构成 $[m,M]$ 的开覆盖,而由 §1 定理 6,必有一个有限子覆盖 $\{U_1,\cdots,U_N,A^c\}$,我们一定可以把 $A^c$ 取走,这不会有影响.于是 $\{U_1,\cdots,U_N\}$ 成为 $A$ 的覆盖,从而 $A$ 为紧.

再证另一方面.设 $A$ 为 $\mathbf{R}^n\,(n=1)$ 的紧子集,今证 $A$ 是 $\mathbf{R}^n\,(n=1)$ 的有界闭子集.因为 $\mathbf{R}^n$ 是豪斯多夫空间,由定理 8 知 $A$ 为闭集,余下的只要证明 $A$ 为有界的即可.作开区间 $I_n=(-n,n)$,则 $\{I_n\}$ 是 $A$ 的一个开覆盖(注意 $A$ 中的元都是数,而 $\pm\infty$ 不是数,所以 $A$ 中任一个 $x$

必位于某个开区间 $I_n$ 中）. 由 $A$ 之紧性，必有有限多个 $I_{n_1},\cdots,I_{n_N}$ 使 $A\subset\bigcup_{k=1}^{N}I_{n_k}$，令 $M=\max_{1\leqslant k\leqslant N}n_k<\pm\infty$，则 $A\subset[-M,M]$ 中从而 $A$ 有界，证毕.

至此，我们已成功地刻画了 $\mathbf{R}^n$ 中的全部紧子集. 那么读者会问，能否类似地刻画度量空间中的全部紧子集即全部有界闭集，至少是刻画巴拿赫空间或希尔伯特空间的全部紧子集即为有界闭集呢？不行，而且这是有深刻的理由的. 所以我们暂时把它放在一边而来进一步讨论紧性. 不过从现在起我们要特别注意一般的拓扑空间与度量空间的区别了. 至少，所有度量空间都是豪斯多夫空间（请读者自己证明），而一般拓扑空间则不一定是.

先看定义在一紧集 $A$ 上的连续函数的性质，这个连续函数 $f(x)$，$x\in A$，可以看成是由 $A\subset X$（$X$ 是一个拓扑空间）到 $\mathbf{R}$ 的连续映射.

**定理 11**　设 $(X,\tau)$ 是一个拓扑空间，$A\subset X$ 是 $X$ 的紧子集. $f(x):A\to\mathbf{R}$ 是 $A$ 上的连续函数，这时 $f(x)$ 必定是有界函数，而且可以达到其上、下确界.

**证**　这个定理就是 §1 定理 7 的 (i)，我们换一个更简单的证法，可以更清楚地看到它与度量性质没有关系. 可以说它是一个纯粹紧性的证明.

由定理 6，$f(A)$ 必为 $\mathbf{R}$ 之紧子集. 由定理 10 又知 $f(A)$ 是有界闭集. 既然 $f(A)$ 作为 $f$ 之值域是有界集，故 $f(x)$ 在 $A$ 上有上、下确界 $M=\sup\limits_{A}f(x)$，$m=\inf\limits_{A}f(x)$，又因 $f(A)$ 是闭集，其上、下确界必是 $f(A)$ 中之点，因此存在 $x_1,x_2\in A$ 使 $f(x_1)=M$，$f(x_2)=m$. 定理证毕.

读者会问，§1 定理 7 之 (ii) 即关于一致连续性又如何？一致连续性是一个度量性质，因此 (ii) 在一般拓扑空间上的推广要利用一种介乎度量结构与拓扑结构中的概念，我们就不再讨论了.

我们在微积分教材中都会看到一个基本定理，即任一无穷集都至少有一个极限点仍在此空间中. 若一拓扑空间有此性质，就说它具有魏尔斯特拉斯 - 波尔查诺性质. 要注意，由定义 6，极限点之定义只与邻域有关. 它是一个一般的拓扑性质，而与序列概念无关，$x$ 是极限点即指 $x$ 之任一邻域中有集合 $A$ 中与 $x$ 不同的点. 这与存在一个收敛于 $x$ 的子序列的概念是两回事. 这一点必须十分注意.

**定理 12**　紧空间 $X$ 必具有魏尔斯特拉斯 - 波尔查诺性质.

**证**　设 $S\subset X$ 没有极限点，今证 $S$ 必为有限集.

任取一点 $x\in X$，必可找到一个开邻域 $U_x$，使得

$$U_x\cap S=\begin{cases}\varnothing, & \text{若 } x\notin S,\\ \{x\}, & \text{若 } x\in S,\end{cases}\tag{6}$$

否则的话，$x$ 将是 $S$ 的极限点. $\{U_x\}_{x\in X}$ 是 $X$ 的一个开覆盖，因为 $X$ 是紧的，故从 $\{U_x\}$ 中可以取出有限多个 $\{U_1,U_2,\cdots,U_N\}$ 来覆盖 $S$，但由 (6) 式，$U_k$ 中或者没有 $S$ 之元，或者只有一个元. 所以 $S$ 必为有限集.

读者会问，这个定理之逆是否成立？如果成立，则我们可以就用魏尔斯特拉斯 - 波尔查诺性质作为紧性的定义，可以明确指出，这是不可能的. 由它只能得到较弱的结论：从 $X$ 之任一可数开覆盖中可以找到有限子覆盖. 因此，紧性的定义只能是定义 12 而不能用“无穷子集必有极限点”来代替，更不能用“任意序列均有收敛子序列”来代替.

下面我们转到度量空间中的紧性. 为此，我们先要给出一个技术性的结果.

**定理 13**(**勒贝格引理**) 设 $X$ 为一紧度量空间,$\{U_\lambda\}$为其一个开覆盖,则存在一个常数 $\delta>0$(称为$\{U_\lambda\}$的勒贝格数),使 $X$ 之任一直径小于$\delta$ 的集合必含于$\{U_\lambda\}$之某一元内.

**证** 用 $\rho$ 表示 $X$ 中的度量,$B_r(P)$表示以 $P$ 为心,$r$ 为半径的球. 一个集合 $U$ 之直径即 $\sup\limits_{x,y\in U}\rho(x,y)$.

我们用反证法证明它. 设定理结论不真,则必存在 $X$ 之子集序列 $A_n$,使所有 $A_n$ 均不含于$\{U_\lambda\}$之任一元 $U_\lambda$ 内,但 $\operatorname{diam} A_n\to 0$. 在 $A_n$ 中选一点 $x_n$,于是得一无穷子集$\{x_n\}$,由魏尔斯特拉斯–波尔查诺性质,这个子集有极限点 $P$,因为$\{U_\lambda\}$是 $X$ 的一个开覆盖,故必有一个 $U\in\{U_\lambda\}$使 $P\in U$. 今取 $\varepsilon>0$,使 $B_\varepsilon(P)\subset U$,再取 $N$ 足够大使得 (i) $\operatorname{diam} A_N<\varepsilon/2$; (ii) $x_N\in B_{\frac{\varepsilon}{2}}(P)$,于是 $\rho(x_N,P)<\dfrac{\varepsilon}{2}$,而对 $A_n$ 之任何一点 $x$,$\rho(x,P)\leqslant\rho(x,x_N)+\rho(x_N,P)<\varepsilon$. 这表明 $A_N\subset U\in\{U_\lambda\}$,这与 $A_N$ 不含于$\{U_\lambda\}$之任一元之内相矛盾.

**注** 以上说$\{x_n\}$是一无穷子集需要解释,因为一个集合中每个元只能计一次(而序列则不然,若说$\{x_n\}$是一序列,即定义在自然数集上的函数,则同一个 $x_n$ 可以出现多次,例如§2 讲到度量空间完备化时讲到的常驻序列),而这里的 $x_n$ 取自 $A_n$ 却是可以重复的. 若 $x_{n_0}$ 重复了无穷多次,我们即可用 $x_{n_0}$ 代替极限点,也就取 $x_{n_0}$ 为 $x_N$,若 $x_{n_0}$ 只重复有限次,则只计一次还会留下无穷多个不同的 $x_n$ 而上面的讨论可以顺利进行.

现在我们就来讨论度量空间的紧性,注意,它与一般拓扑空间是有区别的.

首先来看一致连续性,设 $X,Y$ 是两个度量空间,其度量分别是 $\rho$ 与 $\sigma$,我们说映射 $f:X\to Y$ 是一致连续的(或者准确一点说是关于$(\rho,\sigma)$一致连续的)是指对任意 $\varepsilon>0$,必可找到常数 $\delta(\varepsilon)>0$,使当 $\rho(x_1,x_2)<\delta$ 时,$\sigma(f(x_1),f(x_2))<\varepsilon$,$\delta(\varepsilon)$与 $X$ 中取 $x_1,x_2$之取法无关.

**定理 14** 若 $X,Y$ 为度量空间,其度量为 $\rho$ 和 $\sigma$,若 $X$ 为紧,则一切连续映射 $f:X\to Y$ 必在 $X$ 上一致连续.

**证** 因为 $f$ 已设为连续,故对任给的 $\varepsilon>0$ 以及 $x\in X$ 必可找到 $\tilde{\delta}(\varepsilon,x)>0$,使当 $\tilde{x}\in B_{\tilde{\delta}}(x)$时,$\sigma(f(\tilde{x}),f(x))<\dfrac{\varepsilon}{2}$. 球$\{B_{\tilde{\delta}}(x)\}$构成 $X$ 的开覆盖,因此有一个有限子覆盖$\{B_{\tilde{\delta}_k}(x_k)\}$,$k=1,2,\cdots,N$. 相应于这个覆盖,必有其勒贝格数 $\delta(\varepsilon)>0$,它与 $x_k$ 无关. 今设有任意两点 $x'$与$x''$适合$\rho(x',x'')<\delta$,这时 $x'$与 $x''$必同属于同一个$\{B_{\tilde{\delta}}(x_k)\}$,因此

$$\sigma(f(x'),f(x''))\leqslant\sigma(f(x'),f(x_k))+\sigma(f(x_k),f(x''))$$
$$<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

定理证毕.

在许多微积分教本中常以任一无穷序列$\{x_n\}$均有收敛子序列作为紧性的定义. 这样做明显是有问题的,因为前面我们一再提到,把有关极限的问题归结为序列问题一般说来适用于度量空间. 但是有一个重要的例外. 广义函数理论中的那些函数空间一般不是度量空间,但是在第四、五两章中讲到广义函数空间的拓扑性质时,总是使用序列. 这是需要专门研究的. 本书中我们不能去做这种研究,但是应该指出这样做是必需的,而且是可能的. 在紧性概念问题上我们已经有了

三个概念:以覆盖为基础的紧性定义;以魏尔斯特拉斯-波尔查诺性质为基础的紧性定义;现在又有了以子序列收敛为基础的紧性定义,更确切一点,有

**定义 13** 设 $X$ 为一度量空间,若 $X$ 之任一无穷序列 $\{x_n\}$ 都有收敛的子序列,即有子序列 $\{x_{n_k}\}$ 收敛于 $X$ 中某个元,则称 $X$ 为列紧空间(sequentially compact space).

我们有以下的基本定理.

**定理 15** 若 $X$ 为一度量空间,则以下三个命题等价,

(i) $X$ 为紧;

(ii) $X$ 具有魏尔斯特拉斯-波尔查诺性质;

(iii) $X$ 为列紧.

**证** (i)⇒(ii):即定理 12 应用于度量空间.

(ii)⇒(iii):设 $\{x_n\}\subset X$ 为一无穷序列.定理 13 的注中我们强调了,一个序列 $\{x_n\}$ 与子集 $\{x_n\}$ 之区别在于前者允许有相等的 $x_n$,后者则不允许.今设 $\{x_n\}$ 中有无穷多个彼此相等的元,即有 $x_{n_1}=x_{n_2}=\cdots=x_0$,于是子序列 $\{x_{n_k}\}$ 是一个常驻序列,它自然收敛于 $x_0\in X$.若 $\{x_n\}$ 中没有无穷多个相等的元,则可以找到一个子集仍记为 $\{x_n\}$(即将相等的元只留一个以后删去其多余者),它是一个无穷子集.由魏尔斯特拉斯-波尔查诺性质,它必有一个极限点 $x_0\in X$,但 $x_0$ 之任一邻域中均可找到 $\{x_n\}$ 之至少一个与 $x_0$ 不同的(这是由极限点之定义 6 得知的)点.今任取这样一个点 $x_{n_1}$,并令 $\rho(x_{n_1},x_0)=\varepsilon_1>0$,于是在 $B_{\frac{1}{2}\varepsilon_1}(x_0)$ 中又可找到一点 $x_{n_2}$,使 $\rho(x_{n_2},x_0)=\varepsilon_2>0$,且 $\varepsilon_2<\dfrac{\varepsilon_1}{2}$,仿此可得一个序列 $\{x_{n_k}\}$,使 $\rho(x_{n_k},x_0)=\varepsilon_k<\dfrac{1}{2}\varepsilon_{k-1}<\cdots<\dfrac{\varepsilon_1}{2^{k-1}}$,而且 $\varepsilon_k>0$,很明显 $\lim\limits_{k\to\infty}x_{n_k}=x_0$ 而(iii)得证.

(iii)⇒(i):设 $X$ 为列紧而 $\{U_\lambda\}$ 是 $X$ 的一个开覆盖.由勒贝格引理(这个结果是对紧空间证明的.但是读者容易看到,定理 13 之证明对列紧空间也是有效的),这个开覆盖应有勒贝格数 $\varepsilon$,今取 $0<\delta<\varepsilon$ 并证明必有有限多个以 $\delta$ 为半径的球 $\{B_1,\cdots,B_N\}$,使 $\bigcup\limits_{k=1}^{N}B_k\supset X$,这一点可用反证法证明.若它不对,则以某个 $x_1$ 为心,$\delta$ 为半径的球 $B_1=B_\delta(x_1)$ 不能覆盖 $X$,故有 $x_2\in X\setminus B_\delta(x_1)$.再作 $B_2=B_\delta(x_2)$,则 $B_1\cup B_2$ 仍不能覆盖 $X$,故必有 $x_3\in Z\setminus(B_1\cup B_2)$,而 $\rho(x_1,x_3)\geqslant\delta$,$\rho(x_2,x_3)\geqslant\delta$,再作 $B_3=B_\delta(x_3)$,并找到 $x_4$.仿此,可以找到一个无穷序列 $\{x_n\}$,使 $\rho(x_n,x_m)\geqslant\delta$,$n\neq m$.这样一个无穷序列 $\{x_n\}$ 因为其中任两个元距离不小于 $\delta>0$,而它不可能有收敛子序列,这与 $X$ 之列紧性矛盾.定理证毕.

于是我们看到在度量空间中这三个有关紧性的概念是等价的,而在非度量空间则不一定.这就是在一般拓扑空间中三个概念不能混淆的原因.

下面我们进而讨论一个在许多数学分支中都一再出现的有关紧性的问题.即给出一个空间或其一个子集,弄清楚在什么条件下它们是紧的.因为我们现在主要关心的已不再是有关紧性的一般概念,我们将把注意力放在最常见的空间上,首先是赋范线性空间(包括内积空间).这些空间都是有线性结构的,因此也就有维数,而我们很快就发现,维数在这里起了决定作用.

§1 中我们已看到,$\mathbf{R}^1$ 是局部紧的,即任一点都有一个邻域,其闭包为紧.例如 $\mathbf{R}^1$ 上一点 $x_0$

的邻域就可以取为含此点的区间，其闭包就是闭区间$[a,b]$，而 $x_0$ 为其内点，它当然是紧集——即作为 $\mathbf{R}^1$ 之子空间，在子空间拓扑下是紧空间．$\mathbf{R}^1$ 是线性空间，直观地看就是一条平直的直线，其各点的邻域本质上是一样的，或者用数学语言表述，其上的平移变换是一对一的仿射（但不是线性，因为在平移下原点不是不变的）同构与拓扑同胚．所以要证明 $x_0$ 的邻域闭包为紧只需证明 $x=0$ 的邻域闭包$[-a,a]$，$a>0$ 为紧．再以 0 为中心作一个相似变换，它也是线性同构又加上同胚，而且把$[-a,a]$变成 1 维单位球体$[-1,1]$，所以证明 $\mathbf{R}^1$ 为局部紧，只需证明其以 0 为心的单位球体为紧即可，对于 $\mathbf{R}^n$ 也是一样．

$\mathbf{R}^1$ 为局部紧但本身并不为紧，这是非常重要而又自然的．这都是 $\mathbf{R}^1$ 这个空间的“均匀性”造成的．要证明 $\mathbf{R}^1$ 非紧我们建议用以下证法．以整数 $k$ 为心，$\frac{2}{3}$ 为半径作开区间 $I_k=\left(k-\frac{2}{3},k+\frac{2}{3}\right)$，很明显$\{I_k\}$，$k=0,\pm1,\pm2\cdots$是 $\mathbf{R}^1$ 的开覆盖．我们绝不可能从其中找到有限的子覆盖．因为要覆盖 $\mathbf{R}^1$ 就得覆盖所有的整点 $x=k$，$k=0,\pm1,\pm2,\cdots$而每个整点 $x=k$ 恰好只落在一个 $I_k$ 中．所以想要覆盖这些整点以至覆盖整个 $\mathbf{R}^1$，这些 $I_k$ 一个也不能少．这正是 $\mathbf{R}^1$ 的“均匀性”的表现．

同理 $\mathbf{R}^n$ 也都是局部紧而本身非紧．但转到无穷维情况就完全不同了．我们下面只是不严格地举例以说明，严格证明还要多费一些口舌．设 $H$ 是一个实无穷维希尔伯特空间．于是有一个 o.n.系$\{e_1,e_2,\cdots,e_n,\cdots\}$（我们不问它是否可分，因为不去讨论其完备性），它们都位于单位球面 $\|x\|=1$ 上．单位球面当然是一个度量空间，因此我们想利用定理 15，看一下 $\|x\|=1$ 是否列紧．它当然不是，因为序列$\{e_n\}$中两个不相同的元之距离平方为 $\|e_m-e_n\|^2=\|e_m\|^2-2\langle e_m,e_n\rangle+\|e_n\|^2=2$，而不可能有收敛子序列．如果要严格地讲，我们不加证明地给出

**定理**（F.里斯） 赋范线性空间 $B$ 为局部紧的充分必要条件是 $B$ 为有限维空间．

由此可见，在无穷维空间中一个集合为紧，除了有界条件以外，一定要加上其它条件．循着这个思路可以得到许多在各数学分支中十分有用的结果．下面我们来介绍其中可能是最有名的一个：阿斯科里－阿尔泽拉定理（以下称为 A－A 定理）．阿斯科里（G. Ascoli）和阿尔泽拉（C. Arzelá）都是意大利数学家，他们二人都发现了这个定理．这个定理讲的是有界闭区间 $I=[a,b]$ 上的连续函数空间 $C(I)$ 中的集合为紧的充分必要条件．因为 $C(I)$ 是一个度量空间，所以其中的紧与列紧是一致的．在讨论这个问题前，先作两点说明．

所谓 $M$ 列紧，即 $M$ 中任一序列$\{\varphi_n(x)\}$必有 $C(I)$ 意义下的收敛子序列$\{\varphi_{n_k}(x)\}$：$\lim\limits_{k\to\infty}\varphi_{n_k}(x)=\varphi_0(x)$．$C(I)$ 意义下的收敛即一致收敛．因此，$\{\varphi_{n_k}(x)\}$之极限 $\varphi_0(x)$ 仍在 $C(I)$ 中，即仍为 $I$ 上的连续函数，但不能确定是否在 $M$ 中，而只能确定它在 $\overline{M}$ 中，所以实际上我们只能证明 $\overline{M}$ 为列紧．不过由定理 8，豪斯多夫空间——$C(I)$ 也是——中紧子集必为闭，所以应该证明的只是 $\overline{M}$ 为列紧．若一个集合之闭包为紧，则称此集合为预紧（precompact）．所以我们应该证明的是 $M$ 为预紧．

A－A 定理中要用到等度连续（equicontinuous）概念：若 $M$ 为 $C(I)$ 中之无穷子集（有限子集之等度连续性概念没有意义）为等度连续的，是指对任一 $\varepsilon>0$，必有 $\delta(\varepsilon)>0$ 存在，使当 $|x_1-x_2|<\delta(\varepsilon)$ 时，不论 $x_1,x_2$ 在 $I$ 中何处，也不论 $\varphi(x)$ 是 $M$ 之哪一个元，都有 $|\varphi(x_1)-$

$\varphi(x_2)|<\varepsilon$. 这就是说 $\delta(\varepsilon)$ 不仅不依赖于 $x$（这是一致连续性的条件），而且不依赖于 $\varphi(x)$ 之选择. 于是我们给出 A－A 定理如下：

**定理 16（A－A 定理）** $C(I)$ 之子集 $M$ 为预紧的充分必要条件是 $M$ 中的函数有公共的上界且为等度连续.

**证 必要性** 设 $M$ 为预紧. 我们先用一个概念即 $\varepsilon$ 网来表述紧性. 若度量空间 $X$ 为紧，以任一点 $P$ 为心，$\varepsilon$ 为半径作一球 $B_\varepsilon(P)$，则 $\bigcup\limits_{P\in X}B_\varepsilon(P)$ 覆盖 $X$，因此可以找到有限个这样的球 $B_\varepsilon(P_k)$，$k=1,2,\cdots,N$ 使 $\bigcup\limits_{k=1}^{N}B_\varepsilon(P_k)=X$. $\{P_1,\cdots,P_N\}$ 就称为一个 $\varepsilon$ 网，所以若一度量空间为紧，则对任意 $\varepsilon>0$ 均有有限 $\varepsilon$ 网存在（其逆亦真，这一点下面要用到）. 把它应用于 $M$，并用它来证明 $M$ 中之函数均为等度连续的. 我们构造 $M$ 的 $\dfrac{\varepsilon}{3}$ 网，$\{\varphi_1(x),\cdots,\varphi_N(x)\}\subset C(I)$，每个 $\varphi_k(x)$ 均为 $I$ 上之连续函数，故为一致连续. 因此对 $\varepsilon/3$，可以找到 $\delta_k(\varepsilon)$，使当 $|x_1-x_2|<\delta_k(\varepsilon)$ 时 $|\varphi_k(x_1)-\varphi_k(x_2)|<\dfrac{\varepsilon}{3}$. 在有限多个 $\delta_k(\varepsilon)$ 中取 $\delta(\varepsilon)=\min\{\delta_1(\varepsilon),\cdots,\delta_N(\varepsilon)\}$，显然 $\delta(\varepsilon)>0$，于是对任一 $\varphi(x)\in M$，可以找到 $\varepsilon/3$ 网中的 $\varphi_k(x)$，而得到，当 $|x_1-x_2|<\delta(\varepsilon)$ 时

$$\begin{aligned}|\varphi(x_1)-\varphi(x_2)|&\leqslant|\varphi(x_1)-\varphi_k(x_1)|+|\varphi_k(x_1)-\varphi_k(x_2)|+|\varphi_k(x_2)-\varphi(x_2)|\\&<\varepsilon.\end{aligned}$$

因此 $M$ 是等度连续的.

$M$ 中之元有公共的界易证，因为对上述 $\varepsilon/3$ 网，每个 $|\varphi_k(x)|$ 在 $I$ 上均有上界 $C_k$，令 $C=\max\{C_1,\cdots,C_N\}$，则 $C<+\infty$，而 $|\varphi_k(x)|\leqslant C$. 于是对 $M$ 中任一 $\varphi(x)$ 仿上法有

$$|\varphi(x)|\leqslant|\varphi_k(x)|+|\varphi(x)-\varphi_k(x)|\leqslant C+1,$$

可见 $C+1$ 即 $\overline{M}$ 中所有函数绝对值的公共上界.

**充分性** 充分性的证明基本思想是用有限维空间去逼近 $\overline{M}$，应该强调指出，这是在涉及紧性问题时的一个基本方法.

对于一个连续函数 $\varphi(x)\in\overline{M}$，将 $I$ 等分为

$$a=a_0<a_1<\cdots<a_n=b,\ a_k=a_0+kh,\ h=\frac{b-a}{n}.$$

于是得到 $\varphi(x)$ 曲线上的 $n+1$ 个点 $(a_i,\varphi(a_i))$. 以它们为顶点作折线，记其方程为 $y=\varphi^{(n)}(x)$.（请注意，这里的 $\varphi^{(n)}(x)$ 是指具有 $n+1$ 个顶点的折线的方程右方，而不是 $\varphi(x)$ 的 $n$ 阶导数.）从图形上很容易看到，在子区间 $[a_i,a_{i+1}]$ 中若 $\varphi(a_i)\leqslant\varphi(a_{i+1})$ 必有

$$\varphi(a_i)\leqslant\varphi^{(n)}(x)\leqslant\varphi(a_{i+1}),$$

而当 $\varphi(a_i)\geqslant\varphi(a_{i+1})$ 时则有

$$\varphi(a_{i+1})\leqslant\varphi^{(n)}(x)\leqslant\varphi(a_i).$$

总之，当 $x\in[a_i,a_{i+1}]$ 时

$$\varphi(x)-\varphi(a_{i+1})\leqslant\varphi(x)-\varphi^{(n)}(x)\leqslant\varphi(x)-\varphi(a_i),\tag{$7_1$}$$

或

$$\varphi(x)-\varphi(a_i)\leqslant\varphi(x)-\varphi^{(n)}(x)\leqslant\varphi(x)-\varphi(a_{i+1}).\tag{$7_2$}$$

但是 $\varphi(x)$ 是一致连续的,上式中 $|x-a_i|$, $|x-a_{i+1}|\leqslant\frac{1}{n}$,故当 $n$ 充分大时,$(7_1)$,$(7_2)$ 左右两方之项绝对值均小于 $\varepsilon$,因此对 $x\in I$,有

$$\sup_I|\varphi(x)-\varphi^{(n)}(x)|<\varepsilon, \tag{8}$$

而得到 $\overline{M}$ 的一个由折线组成的 $\varepsilon$ 网,记为 $N_1$.

这些折线还是有界的,因为对每条折线 $\varphi^{(n)}(x)$ 而言

$$|\varphi^{(n)}(x)|\leqslant|\varphi(x)-\varphi^{(n)}(x)|+|\varphi(x)|\leqslant\varepsilon+C\leqslant C+1.$$

这里 $\varphi(x)$ 是 $\overline{M}$ 中适合(8)式的曲线,而 $C$ 是 $\overline{M}$ 中元素绝对值的公共上界.

当然,这个 $\varepsilon$ 网不是有限的,但每一条折线均由固定的分点 $x=a_i$ 处折线的纵坐标 $b_i$ 决定.所以每一条折线对应于 $\mathbf{R}^{n+1}$ 中的一点 $(b_0,b_1,\cdots,b_n)$.因为每个 $|b_i|$ 都有界,故对应于上面得到的 $\varepsilon$ 网的 $(b_0,b_1,\cdots,b_n)$ 之集合是 $\mathbf{R}^{n+1}$ 的一个紧集,因而有一个有限 $\varepsilon$ 网,记为 $N_2$.这个新的有限 $\varepsilon$ 网 $N_2$ 之每一点对应于一条折线 $\psi^{(n)}(x)$.取 $N_2$ 中的 $\psi^{(m)}(x)$,即有

$$|\varphi(x)-\psi^{(m)}(x)|\leqslant|\varphi(x)-\varphi^{(n)}(x)|+|\varphi^{(n)}(x)-\psi^{(m)}(x)|<2\varepsilon.$$

所以 $N_2$ 是 $\overline{M}$ 的 $2\varepsilon$ 网.因为 $2\varepsilon$ 是任意数,我们不妨仍记为 $\varepsilon$ 网 $N_2$.

这个 $N_2$ 中之函数绝对值仍有公共上界.

在必要性部分中我们已证明了,度量空间 $X$ 中任一列紧集均有有限 $\varepsilon$ 网.现在要证明其逆:若这个度量空间 $X$ 还是完备的,而闭子集 $\overline{M}$ 又对任意 $\varepsilon$ 有有限的 $\varepsilon$ 网,则此闭子集 $\overline{M}$ 必为列紧的.这两个结论合起来是豪斯多夫证明的一个定理.我们要特别注意其充分性部分中假设了 $X$ 是完备空间.

现在证明其充分性部分,取一串 $\varepsilon_n\downarrow 0$ 再取 $\overline{M}$ 中一个序列 $S_0=\{\varphi_1,\cdots,\varphi_N,\cdots\}$,先作 $\overline{M}$ 之 $\varepsilon_1$ 网,$\{\psi_1^{(1)},\cdots,\psi_N^{(1)},\cdots\}$,以它们为心 $\varepsilon$ 为半径作球 $B_{\varepsilon_1}(\psi_k^{(1)})$,必可覆盖 $\overline{M}$,因此至少有一个这样的球包含了 $\{\varphi_1,\cdots,\varphi_N,\cdots\}$ 之无穷多个元,记为

$$S_1=\{\varphi_1^{(1)},\cdots,\varphi_N^{(1)},\cdots\}.$$

下面只看序列 $S_1$,再作 $\overline{M}$ 的有限 $\varepsilon_2$ 网以及相应的以 $\varepsilon_2$ 为半径的球,于是又可找到 $S_1$ 的一个子序列全含于此球中:

$$S_2=\{\varphi_1^{(2)},\cdots,\varphi_N^{(2)},\cdots\}.$$

仿此可以作出 $S_2$ 之子序列 $S_3$,$S_3$ 的子序列 $S_4$,…,而每一个子序列都含于半径为 $\varepsilon_k(\downarrow 0)$ 的球中.取这些子序列的对角线序列

$$\{\varphi_1^{(1)},\cdots,\varphi_2^{(2)},\cdots,\varphi_N^{(N)},\cdots\},$$

请读者自己证明它是一个柯西序列,而因 $X$ 为完备的,所以它有一个极限 $\varphi\in X$.又因 $\overline{M}$ 是一闭集,所以 $\varphi\in\overline{M}$.这样得知,在 $\overline{M}$ 中任一序列 $S_0$ 都有一个收敛子序列收敛于 $\overline{M}$ 之某个元 $\varphi$.这样 $\overline{M}$ 是列紧的也就是紧的.定理证毕.

与数学中许多重大的定理一样,A-A 定理在 $L^p(I)$ 中也有由里斯证明了的相应的结果:设 $M\subset L^p(I)$,使得

(i) 存在一个常数 $K$,使得一切 $\varphi(x)\in M$ 均有

$$\int_I|\varphi(x)|^p\mathrm{d}x\leqslant K^p,$$

(ii) 对任意 $\varepsilon>0$ 均有一个 $\delta(\varepsilon)>0$ 存在，使得一切 $\varphi(x)\in M$，只要 $|h|<\delta$ 就有

$$\int_I |\varphi(x+h)-\varphi(x)|^p \mathrm{d}x \leqslant \varepsilon^p, \tag{9}$$

这时，$\overline{M}$ 必为 $L^p(I)$ 中的紧集.

很容易看出，(9)就是 $L^p$ 意义下的等度连续性. 它在上面的讨论中起了关键作用. 因此我们自然要问，它在什么条件下可以成立？我们仍以 $C(I)$ 为例说明它. 若 $\varphi(x)$ 不仅连续，而且有一阶连续导数，于是

$$\varphi(x_1)-\varphi(x_2)=\varphi'(\xi)(x_1-x_2), \qquad x_1\leqslant\xi\leqslant x_2.$$

如果 $\sup|\varphi'(\xi)|=K<+\infty$，则对任给的 $\varepsilon$，令 $\delta(\varepsilon)=\dfrac{\varepsilon}{K}$，即知当 $|x_1-x_2|<\delta(\varepsilon)$ 时 $|\varphi(x_1)-\varphi(x_2)|<\varepsilon$，而如果这个 $K$ 不只适合某一个 $\varphi(x)$，而且适用于 $M$，则 $M$ 是等度连续的. 因此，我们不必假设 $M$ 等度连续，而是换一个空间：

$$C^1(I)=\{\varphi(x), \varphi \text{ 与 } \varphi' \text{在}[a,b]\text{上连续}\}.$$

在 $C^1(I)$ 中我们可以定义范数为

$$\|\varphi\|_{C^1}=\sup_I(|\varphi(x)|+|\varphi'(x)|).$$

很容易看出 $C^1(I)$ 成了一个巴拿赫空间(即完备的赋范线性空间)，这个空间中的有界集

$$M=\{\varphi(x)\in C^1(I), \sup(|\varphi(x)|+|\varphi'(x)|)\leqslant K\}$$

中的元，不但其本身有公共的界 $K$，而且一阶导数也有公共的界 $K$，所以是 $C(I)$ (注意，不是 $C^1(I)$ 而是 $C(I)$)中的有公共的界(不妨称为一致有界)的等度连续函数集. 因此在 $C(I)$ 中为预紧. $C^1(I)$ 作为一个集合是 $C(I)$ 的子集合，我们就说 $C^1(I)$ 可以嵌入在 $C(I)$ 中. 同样上面的 $M$ 也嵌入在 $C(I)$ 中，但是 $M$ 在 $C^1(I)$ 只是 $C^1(I)$ 范数下的有界集，而在 $C(I)$ 中则它——或者说是它的嵌入像——是一个预紧集. 我们就说 $M$ 是紧嵌入在 $C(I)$ 中. 所以，为了应用 A-A 定理，我们先不妨把 $M$ 放在 $C^1(I)$ 中，并证明它是 $C^1(I)$ 中的有界集. 就是说先不妨去证明 $\varphi(x)\in M$ 都有一阶连续导数，然后去估计 $\varphi(x)$ 及其导数 $\varphi'(x)$ 并证明 $\sup(|\varphi(x)|+|\varphi'(x)|)\leqslant K$，$K$ 适用于整个 $M$ 而不只是 $M$ 中某一具体函数. 下一步就是证明 $M$ 可以紧嵌入于 $C(I)$ 中，这样从 $M$ 中的一个序列 $\{\varphi_k(x)\}$ 找出一个收敛子序列. 因此 A-A 定理就是一个紧嵌入定理. 我们只讲了一个最简单的例子，但是读者们应注意到，这是数学和数学物理中极重要的方法. 例如在第三章变分学一节中我们讲到某个泛函(例如能量泛函)的极小化序列，它可能是某个索波列夫空间 $H^k(\Omega)$ 中某个集合 $M$ 的序列. 我们不知道它是否会收敛，但不妨证明它在另一个空间 $H^l(\Omega)$ ($l<k$，可能还要附加一些别的条件)中是紧的，即 $H^k(\Omega)$ 的某个集合 $M$ 紧嵌入于 $H^l(\Omega)$ 中，因此这个极小化序列至少在 $H^l(\Omega)$ 中有收敛的子序列. 当然与上面的 $C(I)$ 和 $C^1(I)$ 不同，这里的讨论要以 $L^2$ 理论为基础，但其基本思想是一致的. 它们也有共同的问题：上面我们虽然涉及 $C^1(I)$，我们得到的子序列都只是在 $C(I)$ 中收敛，其极限函数只能在 $C(I)$ 中而不在 $C^1(I)$ 中. 同样，在变分问题中我们得到的极小化序列也只在 $H^l(\Omega)$ 中，而不在 $H^k(\Omega)$ 中. 这里产生的新问题当然又需要其它方法来解决. 但不论如何，我们看到了紧性问题是整个数学中的一个重大问题，而决不只是怎样把微积分的基本概念弄得更清楚的问题.

关于紧性的讨论至此为止.

**4. 连通拓扑空间** 另一个需要认真研究的问题是空间的连通性问题. 这是一个与空间的度

量没有关系的概念.所以下面我们只设 $X$ 为一个一般的拓扑空间.

所谓连通性直观上就是一个几何形体是一个整体.用一种直观而不严格的说法,如果把它分成两部分 $A$ 和 $B$,且 $A$ 和 $B$ 互不重叠(即 $A\cap B=\varnothing$),那么 $A$ 和 $B$ 就应该紧靠在一起.如果 $A$ 和 $B$ 之间的界面是 $S$,$S$ 上之点不属于 $B$ 就属于 $A$.现在设这些点属于 $B$ 而且是 $B$ 的边界点,则它们虽然也是 $A$ 的边界点,但却不属于 $A$,否则 $A$ 与 $B$ 就有了公共点而互相重叠.所以,如果取 $\overline{A}$,把 $S$ 上这些点也包括进去,则这些点成了 $\overline{A}$ 与 $B$ 的公共点,而有 $\overline{A}\cap B\neq\varnothing$,连通性的数学定义即由此而来.

**定义 14** 设 $X$ 为拓扑空间,若对 $X$ 之任意非空子集 $A,B$,使得 $X=A\cup B$ 者,必有 $\overline{A}\cap B\neq\varnothing$,或 $A\cap\overline{B}\neq\varnothing$,则 $X$ 称为连通空间.若 $X$ 之子集 $C$ 在子空间拓扑下是连通空间,就称 $C$ 为 $X$ 之连通集.反之,若 $X$ 有这样的非空子集 $A,B$,使得 $X=A\cup B$,而同时又有 $\overline{A}\cap B=\varnothing$,$A\cap\overline{B}=\varnothing$,就说 $X$ 是不连通的(disconnected).

如果 $X$ 是不连通的,按定义把 $X$ 分解得 $X=A\cup B$,这时不但有 $\overline{A}\cap B=\varnothing$,而且因为 $A\subset\overline{A}$,所以还有 $A\cap B=\varnothing$,而且 $A$ 与 $B$ 均非空集,那么 $A$ 和 $B$ 有什么性质呢?首先 $X=A\cup B=\overline{A}\cup B\subset X$.读者会问 $\overline{A}$ 不是在 $A$ 上再添加一些元素构成的吗?那么应该有 $\overline{A}\cup B\supset X$,为什么 $\overline{A}\cup B\subset X$ 呢?有一个初学者常常忽略的重要问题:当我们考虑一个拓扑空间 $X$ 时,我们认为 $X$ 就是"一切",就是"宇宙".根本无所谓 $X$ 外之点,所以 $\overline{A}\cup B\supset X$ 是"不通"的.用数学语言来说,$\overline{A}$ 比 $A$ 多出的点必是 $A$ 之导集中之点,但所谓一点 $x\in A'$,首先是指 $x\in X$,然后再断言它还是 $A'$中之点.因此 $A'\subset X$,$\overline{A}\subset X$,$\overline{A}\cup B\subset X$.在子空间拓扑中看这个问题就很清楚.设 $X=\mathbf{R}^2=\{(x,y)\}$,$C=\{(x,y),x^2+y^2<1\}$,则半圆

$$D^+=\{(x,y);x^2+y^2<1,y>0\}\subset C.$$

在 $C$ 的子空间拓扑下,$\overline{D}^+=\{(x,y);x^2+y^2<1,y\geqslant 0\}$,而在 $X=\mathbf{R}^2$ 的拓扑下,$\overline{D}^+=\{(x,y),x^2+y^2\leqslant 1,y\geqslant 0\}$.二者确实是不相同的.总之,在考虑 $C$ 的子空间拓扑时,不能考虑 $C$ 外之点,例如上例中单位圆周 $x^2+y^2=1$ 上之点就不在考虑之列.现在再回到上式,有

$$A\cup B=\overline{A}\cup B=X.$$

但因 $\overline{A}\cap B=\varnothing$,所以 $X\setminus B=A=\overline{A}$,而 $A$ 既然等于 $\overline{A}$ 当然为闭所以 $B$ 作 $\overline{A}$ 的余集当然为开.同理,用 $A\cap\overline{B}=\varnothing$ 又有 $A$ 为开.但这样一来,因 $A\cup B=X$,故 $B$ 作为开集 $A$ 之余集又应为闭.同理 $A$ 也应为闭.这样看来,一个不连通空间一定可以分成两部分,每一部分都是既开又闭的非空子集.在讲到拓扑空间的定义时,我们就看到 $\varnothing$ 与 $X$ 都是既开又闭的.现在我们又看到,一个不连通空间总可以分成非空的两部分 $A$ 与 $B$(即 $X=A\cup B$ 且 $A\cap B=\varnothing$)使每一部分都是既开又闭的.其实这个命题之逆亦成立:若 $X=A\cup B$,$A,B$ 非空,但 $A\cap B=\varnothing$(即 $A$ 和 $B$ 没有公共点),而且 $A$ 与 $B$ 都是既开又闭的,这时 $X$ 一定是不连通的.因为 $A$ 既为又开又闭,故 $A=\overline{A}$,而 $A\cap B=\varnothing$,即意味着 $\overline{A}\cap B=\varnothing$.同理 $A\cap\overline{B}=\varnothing$.再由 $X=A\cup B$,即知 $X$ 是不连通的.总之,我们可以把既开又闭集的概念与连通性概念联系起来,而把以上所说的归结为一条定理.

**定理 17** 设 $X$ 为一拓扑空间,则以下命题是等价的.

(i) $X$ 是不连通空间.

(ii) 存在两非空的既开又闭子集 $A$ 与 $B$,使 $X=A\cup B$,但 $A\cap B=\varnothing$.

(iii) 存在 $X$ 的一个既开又闭子集 $A$,它既非空集,又非全空间.

证明当然略去了.

如果已将 $X$ 分解如上，则可能 $A$ 或 $B$ 还可这样分解，一直到最后，得到 $X=\bigcup_{\lambda} A_{\lambda}$（注意 $A_{\lambda}$ 不一定是有限个甚至不一定是可数多个），而 $A_{\lambda}\cap A_{\mu}=\varnothing\ (\lambda\neq\mu)$，$A_{\lambda}$ 为既开又闭的连通子集. 每一个 $A_{\lambda}$ 称为 $X$ 的一个连通分支，$X$ 为连通空间之充要条件即它只有两个平凡的连通分支：空集 $\varnothing$ 与全空间 $X$.

连通性是一个拓扑性质，因为我们有

**定理 18**　设 $f: X\to Y$ 是由拓扑空间 $X$ 到 $Y$ 的连续映射，若 $X$ 是连通的，则 $f(X)$ 也是连通的.

**证**　我们不妨就设 $Y=f(X)$，否则在 $f(X)$ 上引用子空间拓扑也是一拓扑空间，我们就取它为 $Y$. 设 $Y$ 是不连通的，于是 $Y=Y_1\cup Y_2$，这里 $Y_1, Y_2$ 非空且均为开集. 由 $f$ 之连续性，令 $X_1=f^{-1}(Y_1)$，$X_2=f^{-1}(Y_2)$，则 $X_1, X_2$ 也是非空开集，而且因为 $X=f^{-1}(Y_1\cup Y_2)=f^{-1}(Y_1)\cup f^{-1}(Y_2)=X_1\cup X_2$，$X_1\cap X_2=f^{-1}(Y_1)\cap f^{-1}(Y_2)=f^{-1}(Y_1\cap Y_2)=\varnothing$，所以 $X$ 是不连通的. 这是矛盾. 定理证毕.

既然连通性在连续映射下不变，则在同胚映射下也不变，所以连通性是一个拓扑性质.

现在我们要看一个重要的连通性质，即有

**定理 19**　$\mathbf{R}$ 的子空间 $X$ 为连通的必要充分条件是 $X$ 为一区间（包括开、闭、半开……具有无限端点区间在内）.

**证**　我们先要刻画一下什么是区间. 它就是具有以下性质的 $\mathbf{R}$ 的子集 $X$：若 $a, b\in X$，$a<b$，则 $[a,b]\subset X$.

**必要性**　设 $X$ 在 $\mathbf{R}$ 所赋的子空间拓扑下是连通的，今证 $X$ 是一个区间. 为此我们使用反证法，设 $a, b\in X$，且 $a<b$，但 $[a,b]\not\subset X$，于是必有一点 $c\in[a,b]$，但 $c\notin X$，当然 $a<c<b$. 因为 $a, b$ 都在 $X$ 中，于是令 $X_1=(-\infty, c)\cap X$，$X_2=(c,+\infty)\cap X$，显然 $X_1, X_2$ 非空，因为 $a\in X_1$，$b\in X_2$. $X_1\cap X_2=\varnothing$ 是显然的. 而且因为 $c\notin X$，故 $X_1=(-\infty, c)\cap X=(-\infty, c]\cap X$，$X_2=(c,+\infty)\cap X=[c,+\infty)\cap X$，而 $X_1\cup X_2=\mathbf{R}\cap X=X$. $X_1, X_2$ 显然是 $X$ 在子空间拓扑下的开子集，于是 $X=X_1\cup X_2$ 是不连通的，这与假设矛盾.

**充分性**　设 $X$ 为一区间，今证 $X$ 是连通的，为此设 $X=X_1\cup X_2$，$X_1, X_2$ 非空但 $X_1\cap X_2=\varnothing$，我们来证明 $\overline{X}_1\cap X_2\neq\varnothing$ 或 $X_1\cap\overline{X}_2\neq\varnothing$. 由于已设 $X_1, X_2$ 非空，必可取 $a\in X_1$，$b\in X_2$，不妨设 $a\in X_1$，令 $Y=\{d, d\leqslant b, [a,d]\subset X_1\}$，很明显 $Y$ 非空，因为 $a\in Y$，$Y$ 又是有界集，因为 $b$ 就是 $Y$ 的一个上界. 令 $c$ 为 $Y$ 的上确界，于是有两个可能：(i) $c\notin X_1$，这时 $c\in X_2$，同时 $c\in\overline{X}_1$，就是说 $\overline{X}_1\cap X_2\neq\varnothing$，(ii) $c\in X_1$，这时由于 $c$ 是使 $[a,c]\subset X$ 的最大 $c$ 值，所以必存在一串 $c_n\downarrow c$，使 $c_n\notin X_1$，即 $c_n\in X_2$，于是 $c\in\overline{X}_2$. 这样 $\overline{X}_2$ 和 $X_1$ 中有公共元 $c$ 而 $\overline{X}_2\cap X_1\neq\varnothing$，这就证明了 $X$ 为连通的.

一个区间是连通的本是一个十分直观的事. 但若对连通性作了准确的数学定义后，又是一件有待证明的事，而且证明并非那么简单. 但是在花了些力气以后，会得到报偿，因为可以看见微积分中关于连续函数的中间值定理，其实就是一个关于连通性的定理，而且证明十分简单.

**定理 20（中间值定理）**　设 $f: X\to\mathbf{R}$ 是一连续映射，$X$ 是连通的，若 $f(x)$，$x\in X$，可以取 $a, b$ 两值，则对任一 $c: a\leqslant c\leqslant b$，在 $X$ 中至少可以找到一点 $x_0$，使 $f(x_0)=c$.

**证** 由定理 18 和定理 19，$f(X)$必为 $\mathbf{R}$ 之一个区间 $I$. 因为 $a=f(x_1)$，$b=f(x_2)$均在 $I$ 中，故$[a,b]\subset I$ 即在 $f$ 之值域中，也就是说 $c$ 在 $f$ 之值域中，所以存在 $x_0\in X$ 使 $f(x_0)=c$.

许多读者有一个错误印象，以为这是定义在有界闭区间$[a,b]$(即紧子集$[a,b]$)上的连续函数的性质. 这可能是由于通常微积分教本都把这个性质与闭区间上的连续函数的其它几个基本定理放在一起，而定理的陈述也成了"$f(x)$必可取 $f(a)$与 $f(b)$之间的一切值". 如果 $f(x)$不是定义在闭区间$[a,b]$上，$f(a)$，$f(b)$又有什么意义？ §1 定理 8 基本上是仿照通常教材的写法，现在则给出它完全用连通性的表述. 这时 $X$ 不但不一定是闭区间，甚至连 $\mathbf{R}^n$ 上的集合也不需要：只要 $X$ 是连通的拓扑空间即可，而定理的叙述也认为 $f$ 可取其任意两值(不一定是闭区间的端点)之间的一切值. 与此相似，$f$ 可以取零值的定理也可移到这个情况，只要将定理改为：若 $f: X\to\mathbf{R}$ 是连续映射，$X$ 为连通，则只要 $f(x)$在 $X$ 内可以变号，则 $X$ 中必有 $f(x)$之零点.

在讲到连通性的应用之前，我们再补充一些新的连通空间. 定理 19 表明了区间 $I$ 是连通的，考虑由 $I$ 到一拓扑空间 $X$ 之连续映射 $f: I\to X$. 由定理 18，$f(I)$也是连通的. 但是 $f(I)$就是 $X$ 中的一段弧 $x=x(t)$，$t\in I$，$x\in X$. 它可以是开弧、闭弧或半开半闭的，$I$ 也可以是有界的或无界的区间，如果 $X=\mathbf{R}^n$，它可以写为

$$x_i=x_i(t),\quad i=1,2,\cdots,n,t\in I,$$

就是通常的连续空间曲线，于是它们都是连通的. 这些曲线时常称为路径，$t$ 称为其参数，不过在数学中讲到曲线时，我们时常是指映射 $f$，而不是 $f$ 的像. 曲线是 $X$ 中的连通集，曲面又如何？当然，我们通常的曲面是由 $\mathbf{R}^2$ 的区域 $\Omega$(其中的参数例如可以是$(u,v)$)到 $\mathbf{R}^3$ 的映射：

$$x=x(u,v),y=y(u,v),z=z(u,v).$$

如果 $\Omega$ 是连通的，则曲面也是 $\mathbf{R}^3$ 中的连通集，但 $\Omega$ 是否连通？我们甚至可以问，如果 $\Omega$ 是 $\mathbf{R}^2$ 的矩形，它是否连通？直观上看这不是一个问题，但是如果要证明，则可以利用以下的事实：矩形是两个区间之积，而连通空间的积仍是连通的. 这个事实证明起来也还要花一些口舌.

作为中间值定理的应用，我们给出下面的

**例** 奇数次实多项式

$$p(x)=x^n+a_1x^{n-1}+\cdots+a_n,\qquad a_i\ \text{为实数},$$

至少有一个实零点.

为了证明此定理，只要证明对充分大的 $b>0$，

$$p(b)p(-b)<0,$$

于是在$(-b,b)$中至少有一点 $x_0$ 使 $p(x_0)=0$. 现在令$|x|\geqslant\delta>0$，例如取 $\delta=1$，这样

$$p(x)=x^nq(x)=x^n\left(1+\frac{a_1}{x}+\cdots+\frac{a_n}{x^n}\right),$$

这里

$$q(x)\geqslant 1-\frac{|a_1|}{|x|}-\cdots-\frac{|a_n|}{|x|^n}\geqslant 1-\frac{|a_1|+\cdots+|a_n|}{|x|}=1-\frac{A}{|x|},$$

其中$|A|=|a_1|+\cdots+|a_n|$. 于是若

$$|x|\geqslant 1+2A=R,$$

则不但$|x|\geqslant 1$ 得到满足，而且

$$q(x) \geqslant \frac{1}{2}.$$

特别是因为 $x \to \infty$ 时 $q(x) \to 1$，故知当 $|x| \geqslant R$ 时 $q(x)>0$. 于是由

$$p(b)=b^n q(b), \quad p(-b)=(-1)^n b^n q(-b)$$

知道 $p(b)$ 与 $p(-b)$ 异号，而此定理得证.

由代数学的基本定理，知 $p(x)=0$ 共有 $n$ 个根，但可能是复根. 只不过当 $n$ 为奇数时一定至少有一个实根. 所以 $n$ 为偶数以及其它复根的情况还要用代数学的基本定理来处理. 我们想要提到的是，代数学的基本定理也可以用拓扑学方法去证明.

中间值定理与不动点定理有密切的关系，在§2 中我们已经从度量空间的完备化引申出压缩映像有唯一不动点存在. 在那里，我们把待解的方程 $g(x)=0$ 化为

$$x-f(x)=0, \qquad 即\ f(x)=x-g(x).$$

然后 $g(x)=0$ 之解 $x_0$ 即映射 $f$ 下的不动点：$f(x_0)=x_0$. 现在要讲的不动点原理却与度量空间的完备性无关，我们假设 $f$ 是拓扑空间 $X$ 的子集 $\Omega$ 上的一个连续映射. 我们不要求这个映射是压缩的，但要求 $f$ 把 $\Omega$ 映入 $\Omega$：$f(\Omega) \subset \Omega$. 这种映射可称为自映射. 我们指出，若 $\Omega$ 具有某些性质，则其一切连续的自映射都有不动点存在. 例如我们有

**定理 21（一维的布劳威尔（Brouwer）不动点定理）**　设 $I$ 为 $\mathbf{R}$ 的有界闭区间，$f: I \to I$ 是连续映射，则必有至少一点 $x_0 \in I$，为 $f$ 之不动点：

$$f(x_0)=x_0.$$

**证**　考虑 $g(x)=x-f(x)$，若 $I=[a,b]$，先看 $g(a)$ 与 $g(b)$，因为 $f$ 映连通空间 $I \subset \mathbf{R}$ 到 $f(I)$，所以 $f(I)$ 也是一个区间 $(\alpha,\beta) \subset [a,b]$（也可能是 $[\alpha,\beta)$，$(\alpha,\beta]$ 或 $[\alpha,\beta]$），从而 $\alpha \geqslant a$，$\beta \leqslant b$. 若 $\alpha=a$ 或 $\beta=b$，则 $x=a$ 或 $x=b$ 即为所求不动点. 若 $\alpha=a$，$\beta=b$ 均不成立，则必有 $\alpha>a$，$\beta<b$，因此 $g(a)=a-f(a)=a-\alpha<0$，$g(b)=b-f(b)=b-\beta>0$. 由中间值定理即知一点存在 $x_0 \in [a,b]$，使 $g(x_0)=0$，亦即

$$f(x_0)=x_0.$$

定理证毕.

布劳威尔定理在一个方面不如压缩映像原理. 因为它没有给出例如逐步逼近法那样的实际找出不动点的程序. 它是一个非构造性的存在定理. 同时，它也不给出唯一性. 但是它对映射 $f$ 的要求很少：只要是连续的自映射即可. 与此相对比，它对空间 $X$ 的子集 $\Omega$（现在是区间 $I$）的要求很高. 这里的 $I$ 是 $\mathbf{R}$ 的紧集，但是如果 $X=\mathbf{R}^n$ 是有限维的，紧集并不起重要作用，下面就是一个例子. 设 $\Omega$ 是平面上的圆周，不妨就说是 $S^1$：$x_1^2+x_2^2=1$. 如果把此平面看作是复平面，则 $S^1$ 可以写为 $z=e^{i\varphi}$，令 $f$ 为一旋转：$f: S^1 \to S^1$，$e^{i\varphi} \to e^{i(\varphi+\alpha)}$，$\alpha$ 是常数，很明显，$f$ 是一个连续的自映射，但是它明显没有不动点. 可见是否存在不动点，与 $\Omega$ 的拓扑性质密切相关. 我们将在附录中给出一个一般的布劳威尔定理一个证明：

**定理 22（布劳威尔定理）**　设 $D^n \subset \mathbf{R}^n$ 是 $n$ 维闭球体 $x_1^2+\cdots+x_n^2 \leqslant 1$，$f: D^n \to D^n$ 是一个连续映射，这时 $f$ 在 $D^n$ 中至少有一个不动点.

还要指出，若将 $D^n$ 换成 $D^n$ 中的凸体，这个定理仍成立.

布劳威尔定理本质上属于另一个领域. 这个领域中的问题例如有：某个区域上连续向量场有

没有奇点问题，映射度问题等等.它甚至可以推广到无穷维情况，而且与天体力学、流体力学有密切关系.这个事实说明，拓扑学所涉及的决不仅是微积分的基本概念.从历史上来看，它来自物理学的问题，而且现在更是研究物理世界的有力工具，它为这种研究提供很重要的新观点与新方法.从整个数学发展来看，一度显得很抽象，似乎只是为了逻辑的需要的那些数学理论，后来又越来越直接地可以应用于物理世界的研究，微积分的理论基础涉及的拓扑学理论是一个例子.

# 附录　布劳威尔不动点定理的初等证明

下述的布劳威尔定理是一个重要的拓扑定理：

**布劳威尔定理**　设 $f: D^n \to D^n$ 是一个映 $n$ 维闭单位球体

$$D^n: \quad x_1^2 + \cdots + x_n^2 = \| x \|^2 \leqslant 1 \tag{1}$$

到其自身的连续映射，则 $f$ 必有至少一个不动点 $\bar{x} \in D^n$：

$$f(\bar{x}) = \bar{x}.$$

这个定理涉及拓扑学中许多重要的概念.下面我们将要给出它的一个解析的证明.比之它的拓扑本性，这个证明可以说是初等的.这个证明是米尔诺(J. Milnor)于1978年给出的，原文是：

J. Milnor，Analytic Proofs of the Hairy Ball Theorem and the Brouwer Fixed Point Theorem，Amer. Math. Monthly，vol. 85(1979)，521～524.

该文首先证明了另一个更强的定理：偶数维球面上不可能把头发梳顺，即所谓“头发定理”(hairy ball theorem)：

**定理1**　令 $S^{2n}$ 是 $\mathbf{R}^{2n+1}$ 中的单位球面

$$S^{2n}: \quad x_1^2 + \cdots + x_{2n+1}^2 = 1, \tag{2}$$

则 $S^{2n}$ 上任一个连续切向量场必有零点.

**证明的准备**　我们只需就 $C^1$ 切向量场证明本定理即可，因为任意连续切向量场均可用 $C^1$ 切向量场去逼近.事实上，若 $\xi: S^{2n} \to \mathbf{R}^{2n+1}$ 是一个连续切向量场：$\xi = (\xi_1, \cdots, \xi_{2n+1})$.由魏尔斯特拉斯定理，一定可以找到 $C^1$ 向量 $\xi_\varepsilon(x) = (\xi_1^\varepsilon(x), \cdots, \xi_{2n+1}^\varepsilon(x))$.使得在 $S^{2n}$ 上 $\| \xi - \xi_\varepsilon \|^2 = \sup\limits_{x \in S^{2n}} \sum\limits_{i=1}^{2n+1} | \xi_i(x) - \xi_i^\varepsilon(x) |^2 < \varepsilon^2$.但是 $\xi_\varepsilon(n)$ 不一定是切向量，所以我们把它投影到 $S^{2n}$ 在 $x$ 点的切平面上，得到一个切向量 $\eta_\varepsilon(x)$.很明显 $\eta_\varepsilon(x)$ 仍为 $C^1$ 切向量场，而且仍有 $\| \xi - \eta_\varepsilon \| < \varepsilon$.如果对 $\eta_\varepsilon$ 可以证明定理1，则必有 $x_\varepsilon \in \delta^{2n}$ 适合 $\eta_\varepsilon(x_\varepsilon) = 0$.因为 $S^{2n}$ 是紧的，故当 $\varepsilon \to 0$ 时，可以找到一串 $\varepsilon \to 0$(仍用记号 $\varepsilon$)使 $x_\varepsilon$ 有极限 $\bar{x} \in S^{2n}$，而且

$$\xi(\bar{x}) = \lim_{\varepsilon \to 0} \eta_\varepsilon(\bar{x}) = 0.$$

所以下面我们设定理1中的切向量场 $\xi(x)$ 是 $C^1$ 切向量场.

我们用反证法证明定理1.设 $\xi(x) \neq 0$，于是可以作出 $S^{2n}$ 上一个 $C^1$ 单位切向量 $\nu(x) = \xi(x) | \xi(x) |^{-1}$.这里 $| S(x) |^2 = \sum\limits_{i=1}^{2n+1} \xi_i^2(x)$ 是依赖于 $x$ 的而与上面的记号 $\| S \| = \sup\limits_{x \in S^{2n+1}} | S(x) |$ 不同.以后，我们总是用 $|\cdot|$ 表示一个 $m$ 维向量的欧几里得长度，$m = 1$ 时自然就

是绝对值.有了 $\nu(x)$ 以后我们构造一个单位球面 $S^{2n}$ 的映射

$$\varphi_t(x)=x+t\nu(x),\quad t\in S^{2n}. \tag{3}$$

因为 $x$ 是单位径向量, $\nu(x)$ 是单位切向量,而与 $x$ 正交,故

$$|\varphi_t(x)|=\sqrt{1+t^2}.$$

从而(3)中的 $\varphi_t$ 映单位球面 $S^{2n}$ 到半径为 $\sqrt{1+t^2}$ 的球面(记作 $\sqrt{1+t^2}S^{2n}$)上,再令

$$f_t=|x|\varphi_t(x/|x|) \tag{4}$$

则 $f_t$ 在 $\mathbf{R}_t\times(\mathbf{R}_x^{2n+1}/0)$ 中是一个 $C^1$ 映射,它将半径为 $r$ 的球面 $S_r^{2n}$ 映到 $\sqrt{1+t^2}S_t^{2n}$ 内.

现在我们考虑 $\varphi_t$ 与 $f_t$ 的性质.

**引理 2**　当 $|t|$ 充分小时 $\varphi_t$ 是 $S^{2n}$ 到 $S^{2n}$ 上的微分同胚.

**证**　$\varphi_t$ 显然是 $(t,x)$ 在 $\mathbf{R}_t\times\mathbf{S}^{2n}$ 上的 $C^1$ 映射.但因 $S^{2n}$ 并非欧氏空间,我们只能采取与坐标无关的处理方法.注意到如果限制在 $S^{2n}$ 上考虑,则当 $t=0$ 时,它就是恒等映射,因此其切映射 $\mathrm{d}\varphi_0=\mathrm{id}=I$.而当 $|t|$ 充分小时, $\mathrm{d}\varphi_t$ 也与 $I$ 充分接近而为非退化的,因此可以应用反函数定理知道 $\varphi_t$ 当 $|t|<\varepsilon$($\varepsilon$ 是一适当选定的正数)时是局部微分同胚.但是我们的目的是要证明 $\varphi_t=S^{2n}\to\sqrt{1+t^2}S^{2n}$ 是整体微分同胚.关于可微性不必证明,现在证明它是一个双射.

先证明它是满射.

设 $P\in\sqrt{1+t^2}S^{2n}$ 在 $\varphi_t$ 之像内,则因 $\varphi_t$ 在 $P$ 之某邻域中是微分同胚,所以 $P$ 之某个邻域必全在 $\varphi_t$ 之像中,即是说 $\varphi_t(S^{2n})$ 是一个开集. $\varphi_t(S^{2n})$ 又是一个闭集.因为若 $\{P_k\}\subset\varphi_t(S^{2n})$, $k=1,2,\cdots$,而且 $P_k\to\overline{P}\in\sqrt{1+t^2}S^{2n}$,则必有 $\{Q_k\}\subset S^{2n}$ 使 $P_k=\varphi_t(Q_k)$.因为 $S^{2n}$ 是紧集,由波尔查诺－魏尔斯特拉斯定理知 $\{Q_k\}$ 必有收敛的子序列——不妨设就是 $\{Q_k\}$ 本身,故 $Q_k\to\overline{Q}$,而由 $\varphi_t$ 在 $S^{2n}$ 上的连续性知 $\varphi_t(\overline{Q})=\overline{P}$,因此 $\overline{P}$ 也在 $\varphi_t$ 之像内,因此 $\varphi_t(S^{2n})$ 又是 $\sqrt{1+t^2}S^{2n}$ 之闭集,总之 $\varphi_t(S^{2n})$ 是连通集 $\sqrt{1+t^2}S^{2n}$ 之非空(非空几乎是自明的)又闭又开集,故即为 $\sqrt{1+t^2}S^{2n}$.满射证毕.

再证明 $\varphi_t$ 当 $|t|$ 充分小时是单射,这时要用反证法.设结果不成立,则一定有一串 $t_n\to 0$,而 $\varphi_{t_n}$ 不是单射,即在 $S^{2n}$ 上存在 $\{Q_k\}$, $\{Q_k^1\}$ 使 $Q_k\neq Q_k^1$, $\varphi_{t_n}(Q_k)=\varphi_{t_n}(Q_k^1)$.再利用 $S^{2n}$ 之紧性,可以认为 $Q_k\to\overline{Q}$, $Q_k^1\to\overline{Q}^1$,而 $\varphi_0(\overline{Q})=\varphi_0(\overline{Q}^1)$.但是 $\varphi_0=\mathrm{id}$ 是恒等映射,故 $\overline{Q}=\overline{Q}^1$.这样当 $k$ 充分大时, $Q_k$ 与 $Q_k^1$ 均在 $\overline{Q}$ 的一个邻域中,而 $\varphi_{t_k}$ 在此邻域中是微分同胚,故不可能当 $Q_k\neq Q_k^1$ 时 $\varphi_{t_k}(Q_k)=\varphi_{t_k}(Q_k^1)$.故得单射性质.

引理 2 证毕

在引理 2 的证明中我们避免了使用 $\mathbf{R}^{2n+1}$ 中的坐标.因为在 $S^{2n}$ 上,这 $2n+1$ 个坐标不是互相独立的,但在完成定理 1 的证明时,却要利用 $\mathbf{R}^{2n+1}$ 中的坐标了.

**定理 1 证明的完成**　$S^{2n}$ 是 $D^{2n+1}$ 的表面,现在在 $R^{2n+1}$ 中考虑映射(4),而且就采用 $\mathbf{R}^{2n+1}$ 中的坐标: $x=(x_1,\cdots,x_{2n+1})$, $\nu(x)=(\nu_1,\cdots,\nu_{2n+1})$, $f_t=(f_{t,1},\cdots,f_{t,2n+1})$ 利用(3)有

$$f_t(x)=x+t|x|\nu(x/|x|).$$

所以

$$\det\left(\frac{\partial f_{t_i}(x)}{\partial x_j}\right)=\det\left(\delta_{ij}+t\frac{\partial}{\partial x_j}(|x|\nu_j(x/|x|))\right),\tag{5}$$

而且 $t=0$ 时，这个行列式为1，所以当 $|t|$ 充分小时，这个行列式应该为正.

现在计算 $D^{2n+1}$ 在 $f_t$ 下之像的体积.一方面如果把球体看成一层一层的球壳(其厚度为无穷小)组成，而每一个球壳又看成 $S_r^{2n}$，它在 $f_t$ 映射下成为 $S^{2n}_{\sqrt{1+t^2}r}$，所以

$$\mathrm{vol}(f_tD^{2n+1})=(\sqrt{1+t^2})^{2n+1}\mathrm{vol}D^{2n+1}.\tag{6}$$

另一方面

$$\mathrm{vol}(f_tD^{2n+1})=\int_{D^{2n+1}}\left|\det\left(\frac{\partial f_t}{\partial x}\right)\right|\mathrm{d}x=\int_{D^{2n+1}}\det\left(\frac{\partial f_t}{\partial x}\right)\mathrm{d}x.$$

这里利用了 $\det\left(\frac{\partial f_t}{\partial x}\right)>0$.右方由(5)式应该是 $t$ 的 $2n+1$ 次多项式，但是 $(\sqrt{1+t^2})^{2n+1}$ 不是多项式，这就是矛盾.这个矛盾来自假设 $S^{2n}$ 有一个切向量场 $\xi(x)\neq0$.所以定理1得证.

下面还需要一个定理，它讨论的不是切向量场，而是一般的连续向量场：和定理1的情况不一样，这个定理的结论与空间的维数奇偶性无关，所以我们设空间维数为 $n$.

**定理3**　设 $\xi(x)$ 是 $D^n$ 上的连续向量场.设在 $D^n$ 的边界 $S^{n-1}$ 上，$\xi(x)$ 指向球体 $D^n$ 之外，即在 $|x|=1$ 处

$$\langle x,\xi(x)\rangle\geqslant0.$$

这时 $\xi$ 必在 $D^n$ 中有零点 $\bar{x}$，即 $|\bar{x}|\leqslant1$ 且 $\xi(\bar{x})=0$.

证　分成两种情况

首先设 $n$ 为偶数.把 $\xi$ 拓展为 $2D^n$ (即半径为2的闭球体)上的连续向量场 $\bar{\xi}$ 如下：

$$\bar{\xi}(x)=\begin{cases}\xi(x) & |x|\leqslant1,\\(|x|-1)x+(2-|x|)\xi(x/|x|), & 1\leqslant|x|\leqslant2.\end{cases}\tag{7}$$

很明显 $\bar{\xi}(x)$ 在 $2D^n$ 上是连续的，且在 $1\leqslant|x|\leqslant2$ 处

$$\langle\bar{\xi}(x),x\rangle=(|x|-1)|x|^2+(2-|x|)\langle|x|\xi\left(\frac{x}{|x|}\right),\frac{x}{|x|}\rangle\geqslant(|x|-1)|x|^2>0,$$

这里我们利用了

$$\langle\xi\left(\frac{x}{|x|}\right),\frac{x}{|x|}\rangle\geqslant0.$$

$\left(\text{注意}\frac{x}{|x|}\text{是单位向量}\right)$.所以 $\bar{\xi}$ 在 $2D^n\setminus D^n$ 上不为0.在 $2D^n$ 之边缘即球面 $2S^{n-1}$ 上，因 $|x|=2$，有 $\bar{\xi}(x)=x$ 即是径向向量.

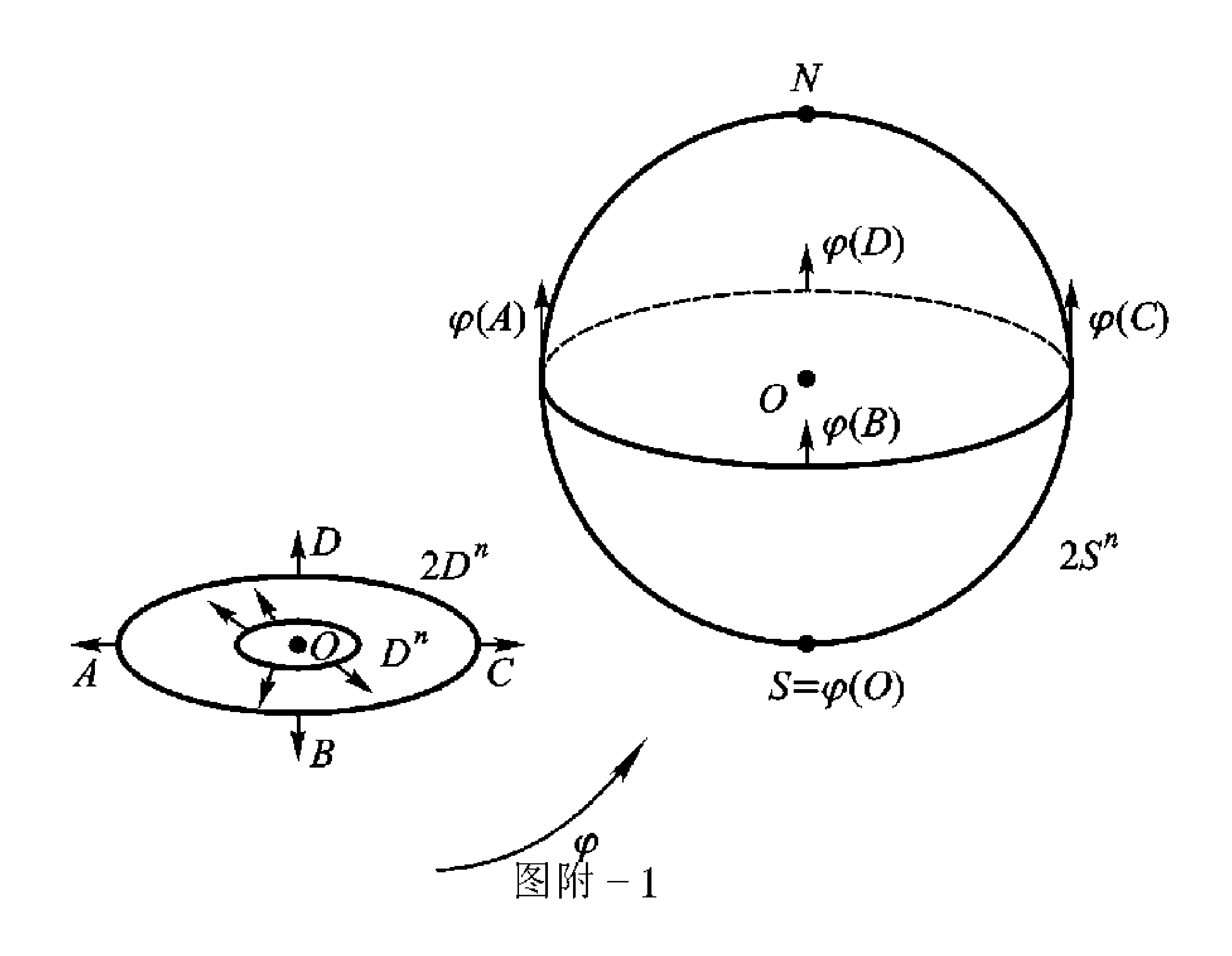
增加一个维数而考虑 $2D^{n+1}$，其边缘是 $2S^n$.现在作一个映射 $\varphi$ 把 $2D^n$ 映到 $2S^n$ 的下半球面上且得其上的切向量场 $\bar{\xi}$，并使在 $2D^n$ 的表面(映为 $2S^n$ 的赤道)上，如 $A$、$B$、$C$、$D$ 四个向量仍映为与赤道平面正交的向量(如图附-1).图附-1只是用几何图形说明它，如果要写出 $\varphi$ 来，则可以利用球极投影.为了避免复杂的运算，我们没有把式子写出来.

现将 $\bar{\xi}$ 拓展到 $2S^n$ 的北半球面成 $\bar{\bar{\xi}}$.如果南半球面上有一点 $P=(x',x_{n+1})$，$x'=(x_1,\cdots,x_n)$，则北半球面必有一个对称点 $Q=(x',-x_{n+1})$.我们定义 $\bar{\bar{\xi}}(Q)$ 之前 $n$ 个分量与

$\bar{\xi}(P)$之前 $n$ 个分量反号，而第 $n+1$ 个分量则同号，换言之

$$\bar{\bar{\xi}}(x',-x_{n+1})=(-x_1,\cdots,-x_n,x_{n+1}).\tag{8}$$

所以可以说 $\bar{\bar{\xi}}$与 $\bar{\xi}$ 是对称的，但因在 $2S^n$ 之赤道上 $\varphi(\xi)$ 与赤道平面正交，故 $\bar{\xi}(x',0)=(0,\cdots,0,\overline{x_{n+1}}),\overline{x_{n+1}}>0$. 由(1)可知 $\bar{\bar{\xi}}(x',0)=\bar{\xi}(x',0)$. 这样得到$2S^n$上的一个连续切向量场，现在 $n$ 是偶数. 由定理 1，拓展后的向量场必有零点，而由对称性在南半球面上至少有一个零点. 由图，零向量在 $\varphi$ 下之原像仍为零向量，而知必有 $\bar{x}\in 2D^n$ 使$\bar{\xi}(\bar{x})=0$，但前已提过 $\bar{\xi}$ 在 $1<|x|<2$ 处不为 0，故 $\bar{x}$ 必在$D^n$中，定理得证.




图附－1

其次看 $n$ 为奇数的情况. 仍与前面相似，我们把维数增加 1 而考虑 $D^{n+1}$，并设法在 $D^{n+1}$上作一个 $n+1$ 维向量 $\bar{\xi}$ 如下：

$$\bar{\xi}(x_1,\cdots,x_{n+1})=(\bar{\xi}(x_1,\cdots,x_n),ax_{n+1}),\tag{9}$$

$a>0$ 待定. 关键在于我们希望拓展后的 $\bar{\xi}$ 在$D^{n+1}$的边缘 $S^n$ 上仍指向$D^{n+1}$之外，即希望在 $x_1^2+\cdots+x_{n+1}^2=1$ 处，

$$\begin{aligned}\sum_{i=1}^{n+1}x_i\bar{\xi}_i(x_1,\cdots,x_{n+1})&=\sum_{i=1}^{n}x_i\bar{\xi}_i(x_1,\cdots,x_n)+ax_{n+1}^2\\&=(x,\xi)+a\left(1-\sum_{i=1}^{n}x_i^2\right)\geqslant 0.\end{aligned}\tag{10}$$

$a$ 的选法如下，和定理 1 证明的准备中说的一样，我们可以假设 $\xi(x)$一直到 $D^n$ 的边界外稍远处，即在$|x|\leqslant 1+\rho,\rho>0$ 处都是连续可微的，并记 $\sum_{i=1}^{n}x_i\xi_i(x_1,\cdots,x_n)=F(r,\theta)$，而$(r,\theta)$ 是球坐标. 于是 $F(r,\theta)$ 对 $r$ 在 $0\leqslant r\leqslant 1+\rho$ 中属于$C^1$. 因此

$$\lim_{r\to 1-0}\frac{F(r,\theta)-F(1,\theta)}{r-1}=F'(1,\theta)$$

是 $S^{n-1}$上的连续函数，所以一定可以找到正常数 $M$ 使

$$\frac{F(r,\theta)-F(1,\theta)}{r-1}\leqslant M,\quad M>0.$$

但是 $F(1,\theta)=\sum_{i=1}^{n}x_i\xi_i(x_1,\cdots,x_n)\Big|_{S^{n-1}}\geqslant 0$，故由上式双方乘以非负的 $1-r$ (注意，我们是从球体 $D^n$ 内接近边界 $S^{n-1}$ 的)，有

$$F(r,\theta)-F(1,\theta)\geqslant -M(1-r).$$

从而

$$F(r,\theta)\geqslant -M(1-r)=-\frac{M}{1+r}(1-r^2)\geqslant -k(1-r^2).$$

所以只要取 $a\geqslant k$，即可证得(10)式. 总之我们有 $\sum x_i\bar{\xi}_i(x_1,\cdots,x_{n+1})\geqslant 0$，现在 $n-1$ 是偶数，所以可以对向量场 $\bar{\xi}$ 应用定理前一部分，而知必有 $\bar{x}=(\bar{x}_1,\cdots,\bar{x}_{n+1})\in D^{n+1}$ 使 $\bar{\xi}(\bar{x})=0$，但

$$\bar{\xi}(\bar{x})=(\xi(\bar{x}_1,\cdots,\bar{x}_n),a\bar{x}_{n+1}),$$

所以 $\bar{\xi}(\bar{x})=0$ 就意味着，$\bar{x}_{n+1}=0$，而 $\bar{x}=(\bar{x}_1,\cdots,\bar{x}_n,0)\simeq(\bar{x}_1,\cdots,\bar{x}_n)\in D^n$. 且 $\xi(\bar{x}_1,\cdots,\bar{x}_n)=0$，定理证毕.

**布劳威尔定理的证明**　对定理中讲的映射 $f$，我们作 $\xi=x-f(x)$，$D^n\to\mathbf{R}^n$，于是 $\xi$ 是 $D^n$ 上的一个向量场，但在 $D^n$ 的边界 $S^{n-1}$ 上，

$$\langle x,\xi\rangle=|x|^2-\langle x,f(x)\rangle\geqslant 0.$$

这是因为 $f(x)$，$D^n\to D^n$，故 $|f(x)|\leqslant 1$，而 $|\langle x,f(x)\rangle|\leqslant|x|\cdot|f(x)|$. 在 $D^n$ 的边界上 $|x|=1$，$|f(x)|\leqslant 1$，所以 $\langle x,\xi\rangle\geqslant 1-|f(x)|\geqslant 0$. 由定理 3 即知 $\xi$ 有一个零点 $\bar{x}$：$\xi(\bar{x})=\bar{x}-f(\bar{x})=0$. 从而 $f(\bar{x})=\bar{x}$ 而 $\bar{x}$ 是映射 $f$ 的不动点，证毕.

以上我们是对单位球体证明布劳威尔定理的. 但是实际上，把 $D^n$ 改成 $\mathbf{R}^n$ 中的闭凸体，而对 $f$ 仍要求它是把此凸体映到其自身中的连续映射，定理仍然成立.

# 第七章　微分流形上的微积分

我们生活于其中的宇宙——时间与空间——是什么样？这是从远古以来的哲学家就十分关注的问题.而自从古希腊时代有了作为一门演绎科学的几何学以后，空间就成了几何学研究的对象.哥白尼的日心说固然在解放人类思想上起了无可估量的作用，为后来的伽利略、开普勒开辟了道路，这以后才有了牛顿力学.但是从对空间的了解而言，只不过是把作为线性空间的欧氏空间换成了一个仿射空间(affine space).什么是仿射空间，我们不来叙述其数学定义而只概略地说明一下.任意定一个起点 $P$(我们不说是原点，仿射空间没有原点).如果再给出一点 $Q$ 作为终点，则联结 $PQ$ 两点可以得一向量 $a=\overrightarrow{PQ}$.现在再以 $Q$ 为 $\mathbf{R}^n$ 的原点，则空间的任一点 $T$ 都可以写成 $a+$ ($\mathbf{R}^n$ 的某点)的形状，即首先在以 $Q$ 为原点的空间中用一个向量 $\overrightarrow{QT}=v$ 来表示 $T$，则 $T$ 相对于 $P$ 点可以表示成为 $\overrightarrow{PT}=\overrightarrow{PQ}+\overrightarrow{QT}=a+v$，这样的集合 $a+\mathbf{R}^n$ 就是一个仿射空间.仿射空间的向量是 $a+v$，它不能作加法：若有两个点 $T_1$ 和 $T_2$ 分别对应于 $a+v_1$，$a+v_2$，它们不能相加，因为形式地相加成 $(a+v_1)+(a+v_2)=2a+(v_1+v_2)$，$2a$ 是什么无法解释.但是可以作减法 $(a+v_1)-(a+v_2)=v_1-v_2$，其实就是 $T_1$ 对 $T_2$ 的相对位置.从数学来看，托勒密和哥白尼的区别就在于托勒密认为原点应该放在 $P$(地球)处，而哥白尼则认为应该放在 $Q$(太阳)处.其实空间的性质没有发生变化：它是三维的、均匀的(在宇宙的任何地方，不论是在比萨斜塔上或在猎户座大星云里，物理规律是相同的)，各向同性的(不论是头朝天顶面向东方还是头下脚上做实验，结果都是一样的).我们说空间是 $\mathbf{R}^3$，除了指定一个原点位置以外，就是这样的意思.至于时间更是与空间无关的一维的、均匀的.如果我们确定了一个时间原点，则时间是 $\mathbf{R}^1$.总之，时空就是 $\mathbf{R}_x^3\times\mathbf{R}_t^1$.这里我们完全没有涉及时间的不可逆性，那是整个物理学的另一大篇章.总之，时空是一个空空洞洞的框架.它是被动的，没有物理内容的.这就是牛顿的时空观.几何学就是研究这个框架的学问.我们在中学学的初等几何和后来学的解析几何学就是这门学问的大概.到了18世纪末叶，情况开始有变化.首先应该提到的是高斯，随着人们枉费心机地想去证明平行线公设(或者证明与此等价的三角形三内角之和为 $\pi$)，高斯开始怀疑我们生于斯长于斯的空间究竟是不是欧氏空间.他真的选了三个小山头，想用实际测量看一下三角形三内角和是否确实为 $\pi$.虽然可惜在他的实验条件下得不出任何结论，却标志着对牛顿的时空观的一次冲击.高斯的一个极重要的贡献就在于提出，不要只把一个曲面看成 $\mathbf{R}^3$ 中弯曲的一部分，应该把曲面就看成是一个空间.在这个空间里，没有毕达哥拉斯定理，而有 $\mathrm{d}s^2=E\mathrm{d}u^2+2F\mathrm{d}u\mathrm{d}v+G\mathrm{d}v^2$，这里 $(u,v)$ 是曲面上的点的"坐标"，$E,F,G$ 是 $(u,v)$ 的函数.这些函数决定了这个曲面的性质，特别是曲面的曲率.而且曲面上的三角形(其三边为测地线)之三角之和与 $\pi$ 之差恰好由曲率决定.黎曼进一步发展了高斯的思想.他指出空间不一定是三维的，而可以是 $n$ 维的，在这样的空间里，毕达哥拉斯定理自然是不对的，而有 $\mathrm{d}s^2=\sum g_{ij}\mathrm{d}x^i\mathrm{d}x^j$.$(g_{ij}(x))$ 完全决定了空间的性质.其实黎曼研究空间的性质是为了说明物理学的问题.可是这个工作不是由黎曼完成的.尽管当时就

有人提出，黎曼的空间是弯曲的，造成这种弯曲是由于物质的分布. 可是到了爱因斯坦才说明了引力与$(g_{ij}(x))$的关系. 这样一来，物理学和几何学就融合起来了. 爱因斯坦提出这个理论——广义相对论——大约是1915年. 当时由于实验观测结果不多，广义相对论对人类的思想影响远不如现今之广泛深刻. 因此，弯曲的空间、高维空间常常成为科幻作品或艺术创作的题材. 但是到了现在，大爆炸、黑洞……一连串重大的科学进展使得空间是弯曲的、空间是高维的这样的思想日渐深入人心，面向21世纪可以肯定在这些方面会取得更大的发展.

这与我们学微积分有什么关系呢？是否可以说微积分某些理论、某些章节已经陈腐了，应该随着知识的"大爆炸"，"炸"到九霄云外去呢？是否可以用物理学来代替它呢？看来都不行. 空间虽然是弯曲的，从局部来看却仍然是通常的 $\mathbf{R}^n$（连仿射概念也用不着），所以微积分的基本理论和方法仍然适用. 但是在下面我们将特别注意一些通常教材上注意不到的方法，所以我们假设读者都掌握了初步的线性代数知识. ——其实在第三章中已经开始这样做了，目的是使读者在将来有机会接触到这类最深刻的数学和物理问题时会少一些生疏感. 这些材料的中心是微分流形问题，但还牵涉到一些其它很有用而时常被忽略的问题.

# §1 向量和张量

**1. 逆变向量与协变向量** "向量就是一个有方向有长度的对象". 这种说法很模糊，但是在力学中我们遇到了许多的向量其共同特点是能按平行四边形法则作加法与倍乘. 而且，例如速度、力等等大家都认为是向量的物理量都有分速度、分力等等，并且可以按平行四边形法则合成这个向量. 把这些概念代数化，就得到线性空间的概念，它的运算规则就是平行四边形法则. 但是我们还要回到几何直观. 说一个向量有方向，有长度，必须有一个参考. 对于 $\mathbf{R}^n$，就是要找 $n$ 个线性无关的向量 $e_1,\cdots,e_n$ 作参考，而按平行四边形法则，每个向量都可写为

$$x = x_1e_1 + \cdots + x_ne_n. \tag{1}$$

$(x_1,\cdots,x_n)$是 $x$ 的 $n$ 个坐标，$x_ie_i$ 称为第 $i$ 个分量（有时也就说 $x_i$ 是第 $i$ 个分量，下面我们一般不加区别）. $(e_1,\cdots,e_n)$称为标架(frame)，或者称为线性空间的一个基底(basis)，在物理学中则称为参考系. 如果一个向量代表某个物理实体或物理量，则在换了一个标架后，该物理量的性质不应该变化，但是标志这个向量的这组分量$(x_1,\cdots,x_n)$却会改变. 因此重要的是了解，当标架变化时分量或坐标如何改变. 首先要看标架作了线性变换后，它们如何改变. 具体说，设我们换了一个新标架$(f_1,\cdots,f_n)$，坐标随之变为$(y_1,\cdots,y_n)$，从物理上看，必须有

$$x = x_1e_1 + \cdots + x_ne_n = y_1f_1 + \cdots y_nf_n. \tag{2}$$

规定(2)式成立是一个重要的原则. 因为一个向量无论它是代表一个物理量或者就简单地是一个几何量，它的性质应该与标架的选取无关. $(x_1,\cdots,x_n)$或$(y_1,\cdots,y_n)$只是同一个向量在不同标架下的表示，各个分量的值可以随标架而变化，但必须有(2)这样的关系式才能断言它们确实表示一个几何或物理实体，这个重要的原则在下面讨论物理定律的表示时将一再遇到.

因为我们现在讨论的是均匀的空间，所以标架的变化应该是线性变换，所以我们令

$$f_i = \sum_{j=1}^{n} \lambda_{ij} e_j, i = 1,2,\cdots,n, \tag{3}$$

这里 $\lambda_{ij}$ 是实数(如果讨论复空间时,则不妨规定 $\lambda_{ij}$ 是复数),而且我们设矩阵

$$\Lambda = (\lambda_{ij}) \tag{4}$$

是非奇异的:$\det\Lambda = |\lambda_{ij}| \neq 0$,因为这时也仅在这时才能从(3)把 $e_j$ 解出写为 $f_i$ 的线性组合.我们不妨引进两个"形式向量"$F = {}^t(f_1,\cdots,f_n)$ 与 $E = {}^t(e_1,\cdots,e_n)$,而把(3)形式地写成

$$F = \Lambda E, \qquad (\text{新标架}) = \Lambda(\text{老标架}). \tag{5}$$

说是"形式向量"是因为真正的向量的分量应该是实数或复数,而不是向量 $e_i$ 或 $f_j$,我们这里只是借用矩阵的形式运算法则而已.以(3)代入(2)即有

$$x_1 e_1 + \cdots + x_n e_n = y_1(\lambda_{11} e_1 + \cdots + \lambda_{1n} e_n) + \cdots + y_n(\lambda_{n1} e_1 + \cdots + \lambda_{nn} e_n)$$

$$= \Big(\sum_{i=1}^{n} \lambda_{i1} y_i\Big) e_1 + \cdots + \Big(\sum_{i=1}^{n} \lambda_{in} y_i\Big) e_n .$$

因为 $e_1,\cdots,e_n$ 是线性无关的,所以有

$$x_j = \sum_{i=1}^{n} \lambda_{ij} y_i, j = 1,\cdots,n. \tag{6}$$

如果引入坐标的向量记法(这一次是真向量了):$x = {}^t(x_1,\cdots,x_n)$,$y = {}^t(y_1,\cdots,y_2)$,则(6)可写成

$$x = {}^t\Lambda y \text{ 或 } y = {}^t\Lambda^{-1} x. \tag{7}$$

$$(\text{新坐标}) = {}^t\Lambda^{-1}(\text{老坐标}).$$

以上,左上角的 $t$ 都表示转置.

(5)与(7)形成重要的对比.两个由 $\mathbf{R}^n$ 到 $\mathbf{R}^n$ 的线性变换,若其矩阵互为转置逆,则说这两个线性变换互为逆步的(contragredient).所以标架的变换(5)与坐标的变换(7)互为逆步.因此,向量 $x$ 的上述坐标 $(x_1,\cdots,x_n)$ 或 $(y_1,\cdots,y_2)$ 称为逆变坐标(contravariant coordinates).所谓逆变是相对于标架而言的,因为标架的变换式是(5)而坐标的变换为

$$y = Ax, A = {}^t\Lambda^{-1}. \tag{8}$$

它是与(5)逆步的变换,所以 $x$ 和 $y$ 称为逆变坐标.

在数学中有一个重要的规定,若一个向量(还有以后要讲的张量)等等,只要含有若干个指标例如 $i,j,\alpha,\lambda,\cdots$ 其变化范围为例如 $(1,2,\cdots,n)$,或 $(0,1,\cdots,n-1)$(视空间的维数而定),而当此空间的坐标作变化(7)时,某个指标是逆变的(即与(7)相同而与(5)逆步),这种指标都写在上角,称为逆变指标.若此指标与(7)逆步,与(5)相同,这个指标一定写在下角,称为协变指标.若在一个式子中同一个字母作为指标出现两次,一次是逆变的,一次是协变的,即一下一上,则规定要对此指标求和,而略去求和号 $\sum$,其变化范围视空间中坐标记号的规定而定.这叫做爱因斯坦求和约定.所以(1),(3),(6),…应写作

$$x = x^i e_i,$$

$$f_i = \lambda_i^j e_j,$$

$$x^j = \lambda_i^j y^i,$$

等等,为什么 $\lambda$ 的两个指标要一上一下,后面要讲.

我们在整个第七章中都使用这样的记号.

既然讲到逆变坐标,当然会问有没有协变坐标(covariant coordinates).它的几何意义是什么?逆变坐标的几何意义前面已经讲得很清楚了,就是把一个向量按平行四边形法则分解在 $n$ 个线性无关的向量(即基底)上.在构成基底的每个向量上应该取一个单位,但这决不等于说在空间中已经引入了长度,因为在度量空间中我们已说过什么是长度.至少任意两点 $x,y$ 间都要定义其距离 $\rho(x,y)\geqslant 0$,而现在我们至多也只能在该向量的方向即该坐标轴上定义"长度".其所以还要加上引号是因为坐标轴上给了两个点 $x$ 和 $y$,$\rho(x,y)=x-y$ 可正可负,表明在 $x$ 和 $y$ 之间要区分出起点与终点.这个概念在度量空间中是没有的.在坐标轴上按给定的单位测"长度",实质上是比例线段问题.欧几里得的《几何原本》就把这一点说得很明白.我们就此停下不再追究,只是说明一点:给定逆变坐标只需要在空间中能作向量运算(加法和正负倍数的放大即前面说的倍乘)就够了.现在我们在 $\mathbf{R}^n$ 中引入欧几里得度量(许多书上这时把 $\mathbf{R}^n$ 改写为 $E^n$,$E$ 是欧几里得(Euclid)第一个字母,本书则不这样做).在 $\mathbf{R}^n$ 中取一个标架,这时又会有一个误解,以为这个标架一定是正交的,这是由于我们太习惯了在 $\mathbf{R}^2$ 与 $\mathbf{R}^3$ 的坐标系 $Oxy$ 与 $Oxyz$,而忘记了标架只需线性无关即可,所以人们忘记了在用到 $Oxy$,$Oxyz$ 时应该加上一句:正交坐标系 $Oxy$ 等等(这当然有点多余).反而在不必加什么说明的地方偏要说在 $\mathbf{R}^n$ 上取一个"斜交坐标系".坐标系本来就无所谓斜正,引入欧氏度量后才得到作为特例的正交标架.不过倒要提醒一下,每个坐标轴上的单位长倒都要选用同样的 1,否则连等边三角形的概念都会出毛病.图 7-1-1上我们只画了 $\mathbf{R}^2$,我们考虑一个向量$\overrightarrow{OP}$以及它在两个坐标轴 $Ox^1$,$Ox^2$ 上的投影长 $\xi_1$ 与 $\xi_2$,于是$\overrightarrow{OP}$完全决定了$(\xi_1,\xi_2)$,反之$(\xi_1,\xi_2)$也完全定义了$\overrightarrow{OP}$,所以$(\xi_1,\xi_2)$也是 $P$ 的一种坐标,下面我们证明这就是 $P$ 的一个协变坐标.

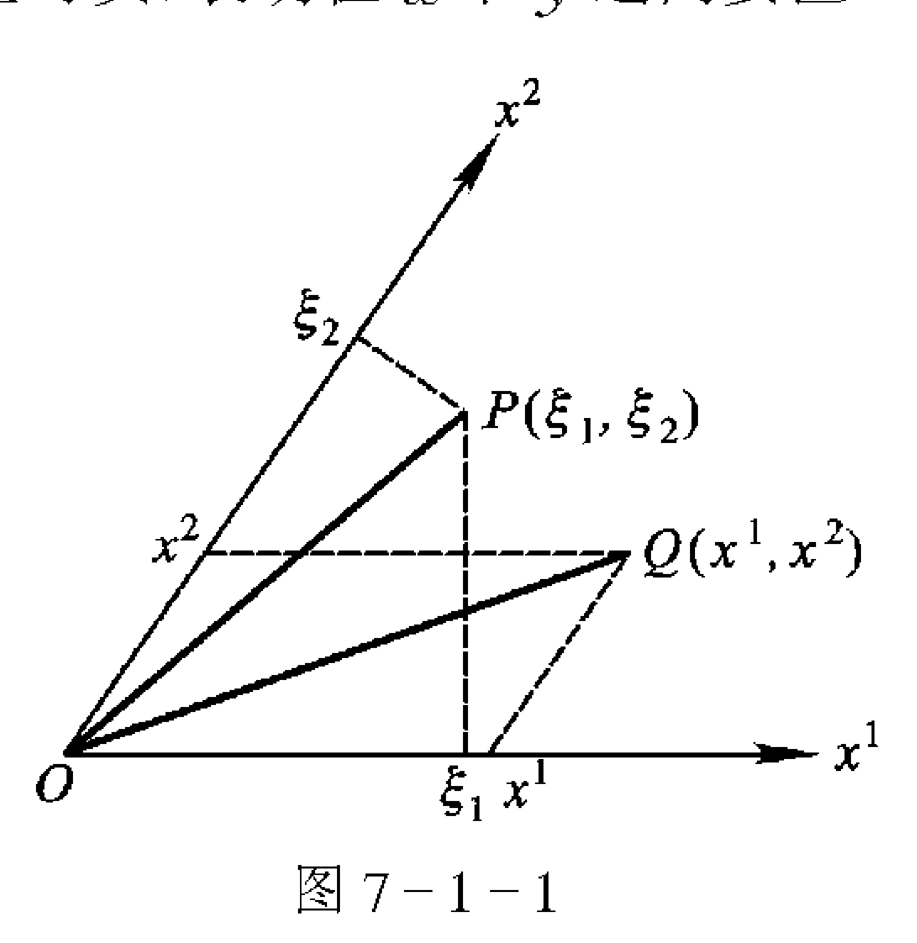


图 7-1-1

为了证明这一点,考虑一个运算:记向量 $A$ 在逆变向量 $B$ 上的投影为 $\langle A,B\rangle$,于是 $\xi_i=\langle\overrightarrow{OP},\boldsymbol{e}_i\rangle$,这个运算对 $A$ 与 $B$ 分别均为线性的,于是任取一个逆变向量$\overrightarrow{OQ}=(x^1,x^2)$;我们有

$$\langle\overrightarrow{OP},\overrightarrow{OQ}\rangle = \xi_1x^1+\xi_2x^2=\xi_ix^i. \tag{9}$$

此式对 $n$ 维向量也是成立的.但是(9)式左方的量显然与坐标的选取无关,所以右方亦然.于是我们换一个标架,使$\overrightarrow{OP}$的投影为$(\eta_1,\eta_2)$,$\overrightarrow{OQ}$的逆变坐标为$(y^1,y^2)$,应有

$$\xi_ix^i=\eta_iy^i. \tag{10}$$

但是新坐标 $y$ 与旧坐标 $x$ 之间有关系式(8),令其中 $A$ 的元素为$(a_j^i)$.$i$ 是第一个指标,表示横行,$j$ 是第二个指标表示竖列.为什么一个写在逆变位置,一个写在协变位置,是因为这样一来,则(8)可以写为

$$y^i=a_j^ix^j. \tag{11}$$

至少可以看到这样写法与 $y^i$ 为逆变坐标相符合.以此代入(10)立即有

$$\xi_ix^i=\eta_ia_j^ix^j,$$

由把上式右的 $j$ 改写成 $i$,$i$ 改写成 $k$,有

$$(\xi_i-\eta_ka_i^k)x^i=0.$$

此式应该对任意 $x^i$ 成立,故其系数为 0,即有

$$\xi_i = a_i^k \eta_k. \tag{12}$$

注意现在是对横行的指标求和而(11)中是对竖列的指标求和,所以(12)右方为 ${}^tA\eta$,而有

$$\eta = {}^tA^{-1}\xi. \tag{13}$$

所以 $\xi_i$ 是协变坐标.

定义逆变坐标时我们只需要 $\mathbf{R}^n$ 的线性结构,也就是只需要平行四边形法则.与逆变坐标不同,在定义协变坐标时,我们则增加了 $\mathbf{R}^n$ 中的欧氏结构——度量,这样才有了投影等等.但是我们想要问,是不是非此不可?不然,我们在上面把投影写成 $\langle A,B\rangle$,读者立刻会想到这是内积,而且就此就联想到欧氏空间的度量.其实我们并没有用到度量的任何特殊性质,如果用 $\langle\xi,x\rangle$,$\langle\eta,y\rangle$ 来表示上面的"内积",我们所用到的只有以下几件事:

1. $\langle\xi,x\rangle = \langle\eta,y\rangle$;
2. $y = Ax$;
3. $\langle\xi,x\rangle = \langle\eta,Ax\rangle = \langle{}^tA\eta,x\rangle$;
4. 由 $\langle\xi-{}^tA\eta,x\rangle = 0$ 对一切 $x$ 成立得出

$$\xi = {}^tA\eta \text{ 或 } \eta = {}^tA^{-1}\xi.$$

第一点是说应该有一个量,此量是内蕴的即与坐标无关的,我们这里则是 $x\in\mathbf{R}^n$ 的一个线性泛函.无限维线性空间的线性泛函之定义中应该加上连续性(即有界性)的要求.我们在讲希尔伯特空间与广义函数时都强调了这一点.但有限维线性空间上的线性泛函则一定是连续的.第四点是要求 $\langle\cdot,\cdot\rangle$ 是非退化的,即若对一切 $x\in\mathbf{R}^n$ 均有 $\langle\xi,x\rangle=0$,则必致 $\xi=0$.在 $\mathbf{R}^n$ 上这总是成立的.至于另外两点就简单地是矩阵的初步知识.在线性代数中我们记 $\mathbf{R}^n$ 上的线性泛函之集合为 $(\mathbf{R}^n)^*$,并且证明了它也是一个 $n$ 维线性空间(当 $\mathbf{R}^n$ 换成 $\mathbf{C}^n$ 时,$(\mathbf{C}^n)^*$ 也是一个 $n$ 维复线性空间)称为 $\mathbf{R}^n$ 之对偶空间,若 $\xi\in(\mathbf{R}^n)^*$,$x\in\mathbf{R}^n$,则 $\xi$ 作用在 $x$ 上之值 $\xi(x)$ 也常用内积记号,记作 $\xi(x) = \langle\xi,x\rangle$,甚至也称为"内积".于是我们可以把以上所说的归结为一个定理.

**定理 1** $\mathbf{R}^n$ 的向量与其对偶空间 $(\mathbf{R}^n)^*$ 中之向量是互为逆步的.因此,$\mathbf{R}^n$ 中的向量称为逆变向量,$(\mathbf{R}^n)^*$ 中之向量称为协变向量.

前面所讲的就是此定理的证明.我们只想提出一点:当我们给出 $\mathbf{R}^n$ 及其对偶空间时,本来都没有给定基底——标架——坐标,所以当我们说 $\xi\in(\mathbf{R}^n)^*$ 在 $x\in\mathbf{R}^n$ 上的值 $\xi(x)$ 或 $\langle\xi,x\rangle$ 时本来就意味着此值与坐标的选择无关.

现在在 $\mathbf{R}^n$ 中取一个基底 $e_i$,$i=1,2,\cdots,n$,注意我们是把指标写在下面的,因为现在我们有了一个标架,而不是向量的坐标——坐标是把向量用 $e_i$ 作线性表达时的系数.当基底作线性变换(3)或(5)时,坐标的变化与之逆步,而这些坐标是用上指标——逆变坐标来表示的,所以标架应该用下指标来表示.前面就是这样做的.与此相同,在 $(\mathbf{R}^n)^*$ 中的标架则应用上指标表示:在 $(\mathbf{R}^n)^*$ 中找一组向量 $(e^*)^j$,$j=1,2,\cdots,n$,使

$$\langle(e^*)^j, e_i\rangle = \delta_{ij}. \tag{14}$$

这样的 $(e^*)^j$ 是存在的,它就是用上式定义的 $\mathbf{R}^n$ 上之线性泛函,很容易看到 $(e^*)^1,\cdots,(e^*)^n$ 是线性无关的,因此构成 $(\mathbf{R}^n)^*$ 的一个基底,称为 $\{e_1,\cdots,e_n\}$ 的对偶基底.

为什么一般的微积分教科书没有区分逆变与协变向量?因为一般书上都只在正交的笛卡儿

坐标系中讨论向量，而两个正交笛卡儿坐标系之间的线性变换又都是正交变换：${}^tA^{-1}=A$，所以(8)和(13)都是一样的，而逆变坐标与协变坐标就没有区别了．我们知道，所有的 $n$ 维实线性空间都是同构的．$(\mathbf{R}^n)^*$ 自然也就同构于 $\mathbf{R}^n$．在丧失了协变与逆变之别后，就更没有理由视它们为不同的空间．如果把 $(\mathbf{R}^n)^*$ 就写成 $\mathbf{R}^n$，把 $(e^*)^i$ 就写成 $e^i$，则对偶基底就成了 o.n. 系．向量 $\xi\in(\mathbf{R}^n)^*$ $(=\mathbf{R}^n)$ 作用在向量 $x\in\mathbf{R}^n$ 上之值，如果用这个 o.n. 系表示就成了

$$\begin{aligned}\langle\xi,x\rangle &= \langle\xi_ie^{*i},x^je_j\rangle = \xi_ix^j\delta_j^i\\ &= \xi_ix^i = \sum_{i=1}^n\xi_ix_i.\end{aligned}$$

也就是内积，即是说，$\mathbf{R}^n$ 的线性泛函，如果选定了一个 o.n. 基底，都可表示成为内积．

$\mathbf{R}^n$ 的一切正交变换均用正交矩阵表示，它构成一个群，记作 $O(n)$．特别重要的是行列式为1的正交矩阵群，是它的一个特别重要的子群 $SO(n)$——特殊正交群．在复空间 $\mathbf{C}^n$ 则有酉群 $U(n)$（物理学家喜欢称为幺正群）和特殊酉群 $SU(n)$．这些都是特别重要的李群（“李”是人名，即 Sophus Lie，1842—1899，挪威大数学家），可以说是现代的理论物理学的最主要的框架．在正交群下，逆变与协变的区别就不起作用了．

但是并不是一切情况下都可以把 $\mathbf{R}^n$ 与 $(\mathbf{R}^n)^*$ 视为相同的．下面我们讲一个极重要的“例子”——$\mathbf{R}^n$ 中的牛顿第二定律．如果我们在 $\mathbf{R}^n$ 与 $(\mathbf{R}^n)^*$ 中采用了互相对偶的基底，而将其中的向量分别写成 $x=(x^1,\cdots,x^n)=x^ie_i$ 与 $\xi=(\xi_1,\cdots,\xi_n)=\xi_j(e^*)^j$，设它们都是时间 $t$ 的光滑函数，于是 $\dot{x}=(\dot{x}^1,\cdots,\dot{x}^n)=\dot{x}^ie_i$，$\dot{\xi}=(\dot{\xi}_1,\cdots,\dot{\xi}_n)=\dot{\xi}_j(e^*)^j$．按我们的习惯，动能是 $T=\sum_{i=1}^n\frac{m_i}{2}\dot{x}^i\dot{x}^i$．但是动能是一个物理量，它应该是与坐标的选取无关的．如果换一个坐标系 $y^i=y^i(t)$，则因 $y^i=\lambda_j^ix^j$，有 $\dot{y}^i=\lambda_j^i\dot{x}^j$，而第 $i$ 个质点之质量 $m_i$ 更是与坐标无关的，所以在 $y$ 坐标中计算的动能是

$$\sum_{i=1}^n\frac{m_i}{2}\dot{y}^i\dot{y}^i = \sum_{i=1}^n\frac{m_i}{2}\lambda_l^i\lambda_k^i\dot{x}^k\dot{x}^l. \tag{15}$$

右方一般说来当然不会是 $T=\sum_{i=1}^n\frac{m_i}{2}\dot{x}^i\dot{x}^i$．但是有一个情况是可以的，即是我们所描述的乃是 $N$ 个质点所成的质点组在 $\mathbf{R}^3$ 中的无约束的运动．这个质点组有 $3N$ 个自由度，例如第一个质点的位置就由 $x_1=X_1$，$x_2=Y_1$，$x_3=Z_1$ 描述，而且 $m_1=m_2=m_3=M_1$ 即第一个质点的质量．这时 $n=3N$ 而且

$$T=\sum_{i=1}^N\frac{M_i}{2}(\dot{X}_i^2+\dot{Y}_i^2+\dot{Z}_i^2).$$

如果作 $\mathbf{R}^3$（注意，不是 $\mathbf{R}^n$）中的正交变换，把 $(X_i,Y_i,Z_i)$ 变成 $(U_i,V_i,W_i)$，这时确实容易计算出 $\frac{M_i}{2}(\dot{X}_i^2+\dot{Y}_i^2+\dot{Z}_i^2)=\frac{M_i}{2}(\dot{U}_i^2+\dot{V}_i^2+\dot{W}_i^2)$，而 $T$ 确实与坐标的选取无关，大多数力学书中都只讨论这个情况．严重的问题在于，作用在这个质点组上的力是保守力，即是有位势为 $F(x)$ 的力，位势就是负位能 $F(x)=-V(x)$，而作用在质点上的力是 $\mathbf{grad}\,F=-\mathbf{grad}\,V$．于是牛顿运动方程成为

$$m_i \ddot{x}^i = \frac{\partial}{\partial x^i} F(x), i = 1, \cdots, n \tag{16}$$

(不对 $i$ 求和). 如果要考虑坐标变换, 立即看出了问题: 式左是逆变向量 $x^i$ 对时间的二阶导数, 在经典力学中时间是一个均匀流动的参数, 它与空间完全无关, 所以本章一开始就说了牛顿时空是 $\mathbf{R}_x^3 \times \mathbf{R}_t^1$ 而不说是 $\mathbf{R}_{(x,t)}^4$, 这样, $x^i$ 对 $t$ 求导后仍是逆变向量. 但式右则不然, 如果引入新坐标.

$$y^i = \lambda_j^i x^j \qquad 或\ y = \Lambda x, \qquad \Lambda = (\lambda_j^i),$$

$$x^i = \mu_j^i y^j \qquad 或\ x = \Lambda^{-1} y, \qquad \Lambda^{-1} = (\mu_j^i),$$

则

$$\frac{\partial F}{\partial y^i} = \frac{\partial F}{\partial x^j}\frac{\partial x^j}{\partial y^i} = \mu_i^j \frac{\partial F}{\partial x^j},$$

$$\frac{\partial F}{\partial x^i} = \frac{\partial F}{\partial y^j}\frac{\partial y^j}{\partial x^i} = \lambda_i^j \frac{\partial F}{\partial y^j}$$

(注意, 分母中的上指标算是下指标). 于是 (16) 的左右双方分别是一逆变与协变向量的分量, 而此式只有在不区分逆变与协变向量的坐标系中, 才有意义. 例如上面说的 $N$ 个质点而每个质点均在一个 $\mathbf{R}^3$ 中运动, 且每一个 $\mathbf{R}^3$ 中各取 o.n. 标架, 就属于这个情况.

怎样克服这里的矛盾? 关键在使动能 $T$ 成为与坐标无关的标量. 方法是在各个坐标系中分别引入一个矩阵——即 $n^2$ 个量 $(g_{ij})$ 并把 $T$ 修改为

$$T = \frac{1}{2} g_{ij} \dot{x}^i \dot{x}^j, \quad g_{ij} = g_{ji}. \tag{17}$$

这里的 $(g_{ij})$ 的选法应该是使在我们原来用以表述牛顿第二定律所用的 $x$ 坐标中 $T = \sum_{i=1}^{n} \frac{m_i}{2} \dot{x}^i \dot{x}^i$, 而在用另一个坐标 $y = Ax, A = (a_j^i)$ 时应把 $g_{ij}$ 改成 $\tilde{g}_{ij}, \tilde{g}_{ij} = \tilde{g}_{ji}$ 使得

$$\tilde{g}_{kl} = g_{ij} a_k^i a_l^j, \quad g_{ij} = \tilde{g}_{kl} \alpha_i^k \alpha_j^l. \tag{18}$$

这里 $(\alpha_i^k)$ 是 $(a_j^i)$ 的逆矩阵. 注意, 我们已经把质量放在 $g_{ij}$ 中去了. 由 (17) 和 (18) 立即有

$$T = \frac{1}{2} g_{ij} \dot{x}^i \dot{x}^j = \frac{1}{2} \tilde{g}_{ij} \dot{y}^i \dot{y}^j.$$

但是 $T$ 既然在 $x$ 坐标系中代表动能, 当然在 $y$ 坐标系中也代表动能——动能是一个标量, 它的值不受坐标变换的影响.

解决了这个问题, 立即就知道, 牛顿第二定律的正确表述应该是

$$\frac{\mathrm{d}}{\mathrm{d}t}(g_{ij} \dot{x}^j) = \frac{\partial F}{\partial x^i}, i = 1, 2, \cdots, n. \tag{19}$$

因为 $\dot{x}^j = \dot{y}^l \frac{\partial x^j}{\partial y^l} = \alpha_l^j \dot{y}^l, \frac{\partial F}{\partial y^k} = \frac{\partial F}{\partial x_i}\frac{\partial x^i}{\partial y^k} = \alpha_k^i \frac{\partial F}{\partial x^i}$ 故将上式双方乘以 $\alpha_k^i$, 再对 $i$ 求和即得

$$\frac{\mathrm{d}}{\mathrm{d}t}(g_{ij} \alpha_k^i \alpha_l^j \dot{y}^l) = \frac{\partial F}{\partial y^k},$$

亦即

$$\frac{\mathrm{d}}{\mathrm{d}t}(\tilde{g}_{kl} \dot{y}^l) = \frac{\partial F}{\partial y^k}. \tag{20}$$

比较(19)和(20)说明牛顿第二定律在任何坐标系中都是成立的,而且可以用简单的形式推导得出各个坐标系下这个定律的形状——现在(19)和(20)形状是一样的——这就叫做这个定律的协变性.一切物理定律都应该具有这种协变性,否则就不能断定它确有物理意义.上面我们是用 $g_{ij}\dot{x}^j$ 代替了速度 $\dot{x}^j$,由于质量已放在 $g_{ij}$ 中,所以 $g_{ij}\dot{x}^j$ 就是动量.但是与速度不同,动量现在是协变向量.因为(19)右方是协变向量,我们用 $\alpha_k^i$ 去乘它再对 $i$ 求和就自然会得到$\dfrac{\partial F}{\partial y^k}$.也就是证明了用 $\alpha_k^i$ 去乘 $g_{ij}\dot{x}^j$ 仍是一个协变向量.这样我们看到,牛顿第二定律是以协变向量形式来表示的.其所以如此做是因为外力是保守力 $\mathbf{grad}F$,而这是一个协变向量.总之,一个有物理意义或几何意义的等式双方必须是同种类型的向量或同为标量,否则就要认真讨论为什么会产生这种不一致.我们上面的讨论与本章一开始就提到的黎曼的思想 $\mathrm{d}s^2 = g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ 有密切关系,$\mathrm{d}s^2$ 既是距离平方,双方除以 $\mathrm{d}t^2$ 即得速度平方 $\left(\dfrac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = g_{ij}\dot{x}^j\dot{x}^j$ 再考虑到可以把质量也吸收到$g_{ij}$中去,则$\dfrac{1}{2}\left(\dfrac{\mathrm{d}s}{\mathrm{d}t}\right)^2$ 其实是动能.所以,黎曼的思想有深厚的物理背景.至此,读者会问,$(g_{ij})$究竟是什么?这是 $n^2$ 个数或函数组成一组(考虑到对称性,实有$\dfrac{1}{2}n(n+1)$个数),而在坐标变换下,它按(18)式变化.它还有一个奇怪的性质:即当它与一个逆变向量 $\dot{x}^j$ 相乘,并对 $j$ 求和,从而使指标 $j$ 在最后的结果中不再出现(这个运算叫"缩并"(contraction)),这时一定出现一个协变向量——动量,我们记作 $p_i = g_{ij}\dot{x}^j$.一个数学对象如果有这些奇妙的性质,其本质必然值得认真讨论.如$(g_{ij})$这样的对象就称为一个张量.黎曼就是发现了度量张量$(g_{ij})$,由此建立了黎曼几何学,对整个数学和物理学起了极大的推动作用,我们将在下面再细说.

最后我们要提出一点,上面我们讲向量都是以空间某一定点为原点建立一个线性空间 $\mathbf{R}^n$,所有的向量都是 $\mathbf{R}^n$ 或$(\mathbf{R}^n)^*$之元,这些空间中的坐标变换也都是常系数的线性变换.但是在实际的物理问题中,并不只讨论一个定点上的问题,而要考虑某一区域 $\Omega$ 中的向量函数、标量函数.这个 $\Omega$ 是 $\mathbf{R}^n$ 的一个区域,但这里的 $\mathbf{R}^n$ 与向量所在的 $\mathbf{R}^n$ 或$(\mathbf{R}^n)^*$是两回事,正如我们的第三章中讲什么是微分时,讲到 $x+h$ 时说的,$x\in\Omega\subset\mathbf{R}^n$ 与$h$ 所在的 $\mathbf{R}^n$ 是不同的.后者叫切空间,前者叫底空间.这在下一节讨论切丛、余切丛时会再讨论.现在我们讲坐标变换时就要考虑 $\Omega$ 中的$x$ 变动,而不只是 $\mathbf{R}^n$ 中的向量按$y=Ax$ 或$y={}^tA^{-1}x$ 的变换,但是现在的非线性变换 $x=x(y)$的线性部分,亦即下面要讲的切变换,确实是以底空间中的 $x$ 点或相应的$y$ 点(均作为定点考虑)为原点的切空间 $\mathbf{R}^n$ 中的线性变换,这个线性变换的矩阵就是雅可比矩阵$\left(\dfrac{\partial x^i}{\partial y^j}\right)$.这样,我们就可以给出

**定义 1** 设在 $n$ 维区域$\Omega$ 中作任意坐标变换,即微分同胚 $x=x(y)$或 $y=y(x)$.若在 $\Omega$ 上的 $n^0$ 个光滑函数 $f(x)$,在 $y$ 坐标中相应也有$n^0$ 个光滑函数 $F(y)$,使

$$f(x) = F(y), \quad x = x(y), \tag{21}$$

就称 $f(x)$为一标量.若在 $\Omega$ 中对每个坐标系有$n^1$ 个光滑函数所成的一组$(f_1(x),\cdots,f_n(x))$,而在另一坐标系 $y$ 下相应有一组$(F_1(y),\cdots,F_n(y))$间有关系式

$$F_i(y) = \sum_j f_j(x)\frac{\partial y^i}{\partial x^j}, \tag{22}$$

则称它为一逆变向量,这时规定将指标写在上面而有

$$F^i(y) = f^j(x)\frac{\partial y^i}{\partial x^j}.$$

若 $f$ 与 $F$ 间的关系式为

$$F_i(y) = f_j(x)\frac{\partial x^j}{\partial y^i},$$

这时规定指标写在下面,则称之为协变向量.

例如,$\Omega$ 中某一函数 $\varphi(x)$ 之梯度 $\mathbf{grad}\ \varphi$ 就是协变向量. $\Omega$ 中某一曲线 $x^i = x^i(t)$ 之"导数"或"速度" $\left(\frac{\mathrm{d}x^1}{\mathrm{d}t}, \cdots, \frac{\mathrm{d}x^n}{\mathrm{d}t}\right)$ 以及沿此曲线的微分 $(\mathrm{d}x^1, \cdots, \mathrm{d}x^n)$ 都是逆变向量. 可以说,它们是最重要的协变和逆变向量.

**2. 旋转和反射**　上一节中除了最后一段是讨论的非线性坐标变换以外,我们都是讨论的线性变换. 在物理学中具有特殊重要性的是一些特殊的线性变换例如 $SO(n)$, $U(n)$ 还有作为相对论基础的洛伦兹变换(Lorentz transformation)等等. 这就涉及了群论的问题,是一个浩瀚的数学海洋. 即使是从微积分的角度来看,也确有一些变换特别重要,例如有平移、旋转和反射. 我们在这里不去讨论平移,因为它虽然简单,却不属于线性变换,而是仿射变换. 下面我们讨论旋转. 至于反射,则其性质又与旋转很不相同. 不过,这里有两个相联系的问题. 前面我们是固定了一个向量,即一个点,然后让基底——即坐标轴——改变并且看此点的坐标如何变化. 另一个情况是固定基底,并让此向量,即一个点,变成另一个向量,而且这个变换对于被变换的向量是线性的. 这两种问题虽然不同,却互有联系,甚至可以看成一种"相对运动":把一个平面向量绕原点旋转角度 $\varphi$ 与将坐标轴旋转一个角 $-\varphi$,所得的公式是一样的. 所以下面我们只看向量的变化.

先看平面上的旋转,先看由 $(x, y)$ 点(现在没有必要区分协变与逆变向量,因为我们用的是正交笛卡儿坐标)到 $(x', y')$ 的旋转,其旋转中心为原点,旋转角为 $\varphi$. 最好的方法是用复数记号,于是令 $z = x + \mathrm{i}y$, $z' = x' + \mathrm{i}y'$,则

$$z' = \mathrm{e}^{\mathrm{i}\varphi}z, \tag{23}$$

$$(x' + \mathrm{i}y') = (x\cos\varphi - y\sin\varphi) + \mathrm{i}(x\sin\varphi + y\cos\varphi),$$

分开实虚部,并用矩阵表示,即得

$$x' = x\cos\varphi - y\sin\varphi, y' = x\sin\varphi + y\cos\varphi,$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{24}$$

这是点的变换. 如果坐标轴旋转一个角度 $\varphi$ 则得

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{24'}$$

不论是哪个问题,变换矩阵都是正交矩阵.

上面我们还看到,平面旋转可以用两个观点来看:一是看成实的正交变换,一是看成复变换 $z' = \mathrm{e}^{\mathrm{i}\varphi}z$, $\mathrm{e}^{\mathrm{i}\varphi}$ 可以看成一个一阶矩阵,它仍是酉矩阵,因此记为 $(\mathrm{e}^{\mathrm{i}\varphi})$. 这个情况在三维时照样成立.

其实(24)或(24′)也可以看成一个三维的点变换或坐标变换. 例如(24′)是 $x$, $y$ 平面绕 $z$ 轴旋转了角度 $\varphi$. 因为 $z$ 轴是旋转轴, 所以 $z$ 坐标不变 $z'=z$. 这样, 例如(24′)就成为

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{25}$$

变换矩阵仍是一个正交矩阵, 不过阶数为 3. 我们要证明 $\mathbf{R}^3$ 中一般的旋转都可以这样处理. 现在设原有一个坐标系 $Oxyz$, 经旋转成为 $Ox'y'z'$. 我们要想办法把一个一般的旋转分解为若干个(25)那种类型的旋转的乘积. 为此, 我们令 $OL$ 为 $Oxy$ 平面和 $Ox'y'$ 平面之交线, $Oz$ 和 $Oz'$ 之交角是 $\theta$, 由 $Ox$ 到 $OL$ (均在平面 $Oxy$ 上)之角度为 $\varphi$, 在 $Ox'y'$ 上由 $Ox'$ 到 $OL$ 之角度为 $\psi$ (图 7-1-2). 由 $Oxyz$ 坐标系变到 $Ox'y'z'$ 分成三步实现.

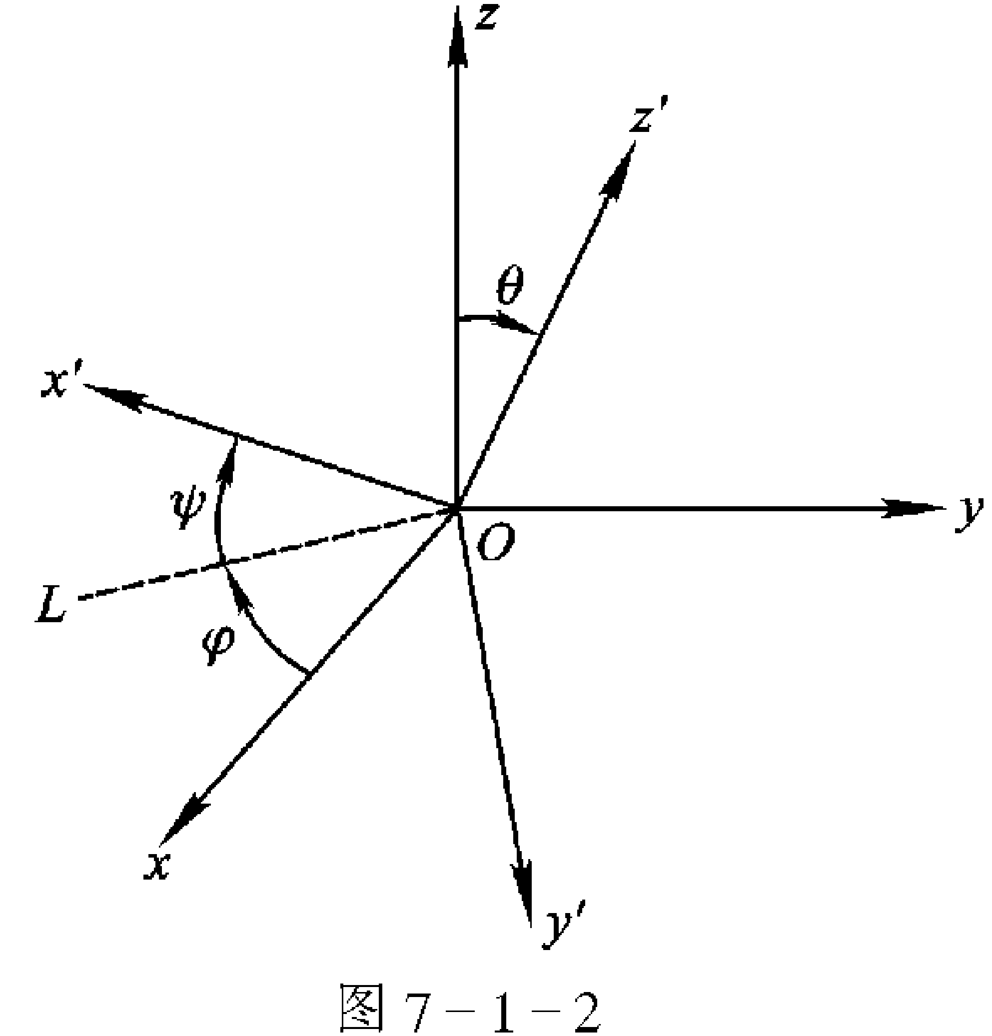


图 7-1-2

1. 以 $Oz$ 为轴, 在 $Oxy$ 平面上将 $Ox$ 旋转 $\varphi$ 成为 $OL$, 令新坐标系为 $(x_1, y_1, z_1)$, 则 $Ox_1$ 即 $OL$, $Oz_1$ 即 $Oz$, 而由(25)有

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{26}$$

2. 现在 $Ox_1$ 即 $OL$, 它既在 $Ox_1y_1$ 平面上, 又在 $Ox'y'$ 平面上, 所以一方面它与 $Oz_1$ 正交, 一方面又与 $Oz'$ 正交. 以 $Ox_1$ 为轴在 $Ozz'$ 面上旋转一个角度 $\theta$, 于是得到新坐标系 $Ox_2y_2z_2$, 这里 $Ox_2$ 即 $Ox_1$, $Oz_2$ 即 $Oz'$ 而由(25)式有

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}. \tag{27}$$

3. 现在绕 $Oz_2$ 作最后一次旋转. 因为 $Oz_2$ 即 $Oz'$ 所以旋转是在 $Ox_2y_2$ 平面上实现的. 旋转的角度为 $\psi$. 于是 $Ox_2$ 即 $OL$ 变成 $Ox'$. $Oz_2$ 是旋转轴所以不变. $Oy_2$ 即 $Oy'$ (条件是 $Ox'y'z'$ 与 $Oxyz$ 同为右手或左手坐标系), 于是最后有

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}. \tag{28}$$

合并(26), (27), (28)即得最后的结果

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{29}$$

$$A = \begin{pmatrix} \cos\varphi\cos\psi - \cos\theta\sin\varphi\sin\psi & \sin\varphi\cos\psi + \cos\theta\cos\varphi\sin\psi & \sin\theta\sin\psi \\ -\cos\varphi\sin\psi - \cos\theta\sin\varphi\cos\psi & -\sin\varphi\sin\psi + \cos\theta\cos\varphi\cos\psi & \sin\theta\cos\psi \\ \sin\varphi\sin\theta & -\cos\varphi\sin\theta & \cos\theta \end{pmatrix}.$$

很容易证明, $A$ 是一个正交矩阵, 而且 $\det A = 1$. 但是要注意, 这是由于我们假设了 $Ox'y'z'$ 与

$Oxyz$ 同为右手或左手系，即有相同"手征"(chirality)，否则下面我们会看到，$A$ 仍为正交矩阵，但是 $\det A=-1$，这种情况将在下面讨论．所有这种适合 $\det A=1$ 的正交矩阵成为一个群，记作 $SO(3)$，$(\theta,\varphi,\psi)$ 是它们的三个参数，称为欧拉角，它是欧拉在研究刚体绕固定点的旋转时提出的．固定点即选为原点 $O$，刚体原来的位置通过附着于其上的以 $O$ 为原点的一个正交标架 $Oxyz$ 来刻画，而其转动后的位置则通过 $Ox'y'z'$ 来刻画．

我们把这些结果归结成为

**定理 2**　绕固定 $O$ 点的正交坐标系的旋转成为行列式为 1 的正交矩阵集合 $SO(3)$．它是一个群，称为三维特殊正交群，它含有三个连续参数 $(\theta,\varphi,\psi)$（称为欧拉角），这里 $0\leqslant\theta<\pi$，$0\leqslant\varphi<2\pi$，$0\leqslant\psi<2\pi$．

上面关于参数变化范围的规定是很明显的．此外，我们还应证明 $SO(3)$ 之任一元一定刻画正交标架的一个旋转，这是很容易证明的，因为正交变换下长度不变，所以角度也不变．这样正交标架仍变为正交标架，行列式为 1 表明标架的手征不会改变．

重要的是，如果我们把 $n\times n$ 矩阵的 $n^2$ 个元看成 $\mathbf{R}^{n^2}$ 中一点的坐标，于是每个 $n\times n$ 矩阵 $A$ 就成了 $\mathbf{R}^{n^2}$ 中的一点．矩阵为非奇异的，即 $\mathbf{R}^{n^2}$ 中不适合代数方程 $\det A=0$ 的点之集合．这是一个开集，这类矩阵也成了一个群，称为 $n$ 维一般线性群，记作 $GL(n)$（有时写作 $GL(n,\mathbf{R})$ 以与复矩阵群 $GL(n,\mathbf{C})$ 区别）．$SO(3)$ 是 $GL(3)$ 的一个子群，而由正交条件（共 6 个方程）来刻画，所以这是一个 3 维曲面．这个曲面是光滑的，具有微分流形结构．每一个点因为代表一个矩阵又可以对它进行群运算．这是一大类极为重要的数学结构，称为李群，我们在这里当然不能再讲了．

上面我们提到，二维平面上的旋转可以用一个复数 $e^{i\varphi}$ 来表示．它也是一个一阶酉矩阵，所以 $SO(2)$ 也就是 $SU(1)$，同样 $\mathbf{R}^3$ 中的旋转 $SO(3)$ 也可以用 $SU(2)$ 表示．这里我们就不来证明了．

需要提到的是行列式为 1 的条件，但实际上，若 $A$ 为正交矩阵，则 ${}^tA=A^{-1}$ 所以，${}^tA\cdot A=I$，从而 $\det({}^tA)\cdot\det(A)=|\det A|^2=1$，而 $A$ 之行列式之值可以是 $\pm1$．上面我们讲了 $A\in SO(3)$ 必得 $\det A=1$，现在要问，$\det A=-1$ 是什么样的标架变换即坐标变换？这是一个有深刻物理意义的问题，因为有两种坐标系，如图 7-1-3 中 $\mathbf{R}^2$ 的两个正交标架就不可能用旋转与平移使之重合．而其中之一在镜子（虚线）中的像就是另一个．这两个互为镜像的坐标系称为具有不同手征：右方的一个称为"右手坐标系"，左方的一个称为"左手坐标系"，二者的关系是

$$\begin{pmatrix}x'\\y'\end{pmatrix}=\begin{pmatrix}-1&0\\0&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.\tag{30}$$

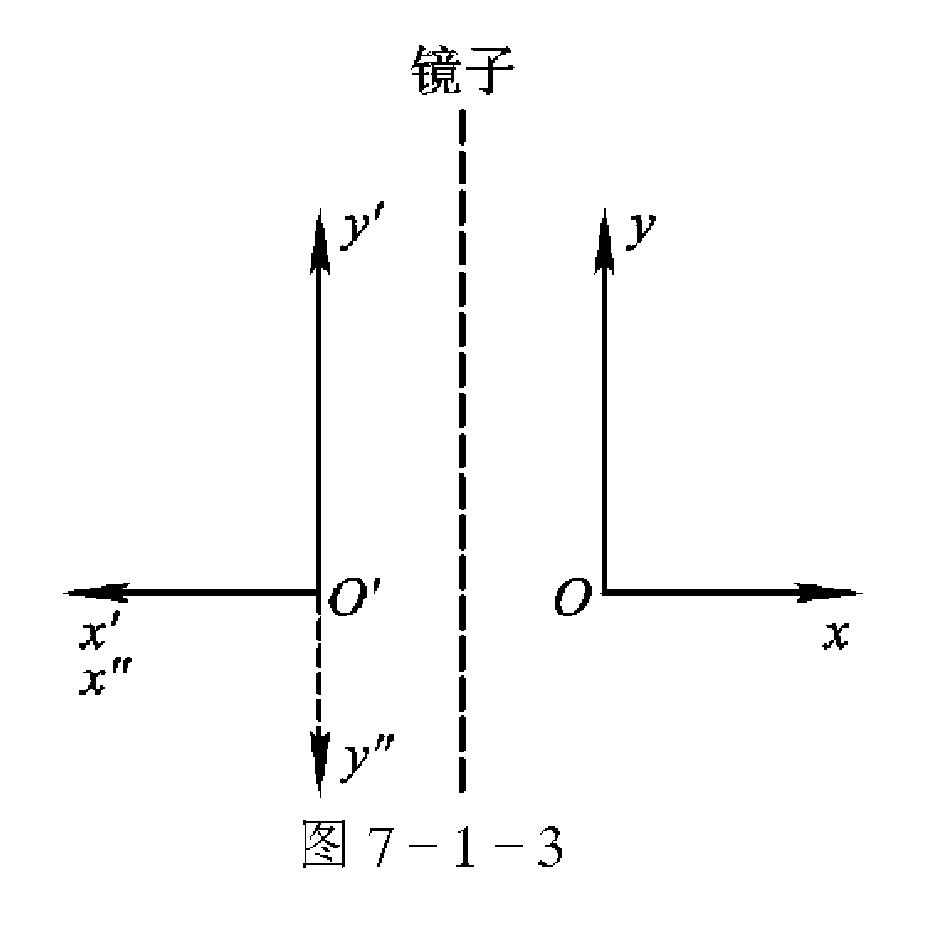


图 7-1-3

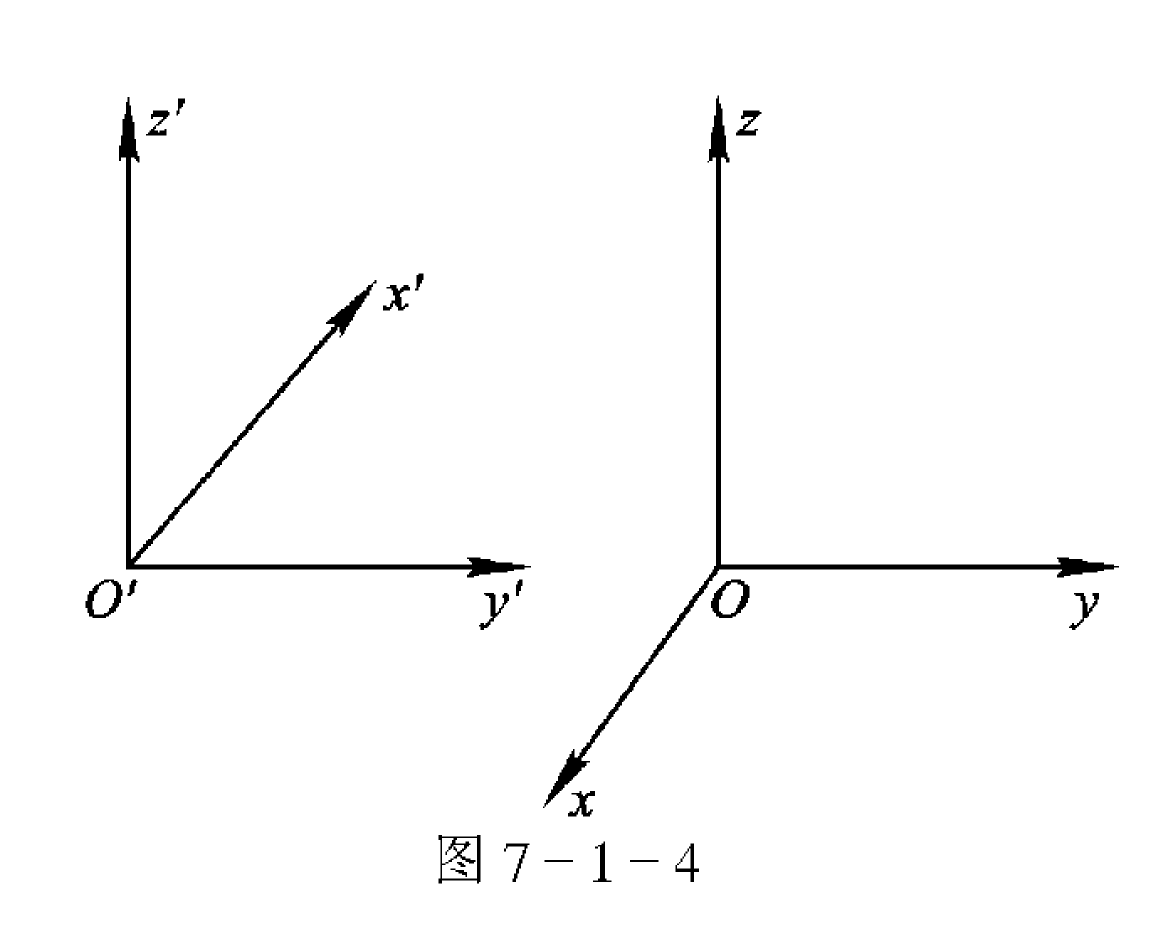


图 7-1-4

对于 $\mathbf{R}^3$ 也是一样，图 7－1－4 给出了左、右手坐标系，其关系也是

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{31}$$

(30)和(31)这样的线性变换称为 $x$ 方向的反射(reflection). 如果对所有坐标轴均作反射，则称为反演(inversion)(物理学家常称为宇称(parity)变换). 反射与旋转不同，它不形成群. 例如在 $\mathbf{R}^2$ 上作对 $x$ 与 $y$ 轴两次反射如图7－1－3的 $O'x''y''$，等于绕 $O$ 点旋转 $\pi$，而不再是反射.

坐标系的手征是一个重要问题. 在 $\mathbf{R}^3$ 中选取一个 o. n. 标架时，即使原点位置相同，三个轴都与对应的轴平行，却可以有手征的不同. 这会影响到对物理量的表述和刻画. 因为物理量和几何量与参考系——坐标系的手征的关系是非常重要的问题. 即以"向量"而言，我们前面指出，所谓的向量原来只是指线性空间之元素，因此适用平行四边形法则，从而也有坐标，而且坐标的变换与标架的变换是逆步的，因而是逆变坐标. 有了对偶空间以后又有了协变坐标. 不过这一段讨论又限制坐标变换只是正交变换，因此协变逆变之区别又消失了. 但在这时又出现了新问题，即正交变换是否保持手征不变；从算术上看即此变换的行列式是 $+1$ 还是 $-1$；从几何上看则是此变换中是否除了旋转以外还包含了反射. 于是，从作为线性空间的元素的向量中又可分为两个子类：一类是在一般的正交变换，即在 $O(3)$之下，亦即容许手征改变时，适合坐标变换的规则，我们甚至说它们在手征变换下不变(或者比较准确一些说它协变)，这类向量称为真向量(proper vector)或极向量(polar vector)，另一类在 $SO(3)$下按坐标变换的规则处理，而若手征改变(例如 $x'_1 = -x$ 之类)时，则整个向量变号，亦即在 $\det A = -1, A \in O(3)$时，坐标变换 $A$ 将写成 $A = PA_1$，$P$ 为手征变换，$Pu = -u$，$A_1 \in SO(3)$按通常的向量变换规则作用于一向量 $u$，于是 $Au = PA_1 u = Pv = -v = -A_1 u$. 这一类向量称为赝向量、伪向量(pseudo－vector)或轴向量(axial vector). 上面讲的话可能令读者感到太抽象，但看一下实例就知道这个区别实际上是很本质的. 在 $\mathbf{R}^3$ 中最重要的真向量当然是位置向量$(x, y, z)$以及它对时间的导数，即速度 $\boldsymbol{v}$ 与加速度 $\boldsymbol{a}$. 由牛顿第二定律得出的力 $\boldsymbol{F} = m\boldsymbol{a}$，还有动量 $m\boldsymbol{v}$ (注意，现在我们不区别协变与逆变向量)，当然也都是真向量. $\mathbf{R}^3$ 中最常见的赝向量是两个真向量 $\boldsymbol{A} = (a_1, a_2, a_3)$与 $\boldsymbol{B} = (b_1, b_2, b_3)$的向量积 $\boldsymbol{A} \times \boldsymbol{B}$. 请读者回忆一下向量积的定义. 在通常的教本中的说法，例如：$\boldsymbol{A} \times \boldsymbol{B}$ 是一个向量，其大小是 $|\boldsymbol{A}||\boldsymbol{B}|\sin\varphi$ ($\varphi$ 是 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 之间的劣角，所以 $0 \leqslant \varphi \leqslant \pi$)，其方向与 $\boldsymbol{A}$、$\boldsymbol{B}$ 所定义的平面垂直 而且 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{A} \times \boldsymbol{B}$ 成一个右手系，这里有一个隐含的假设即 $Oxyz$ 是一个右手系. 如果想丢掉这个假设，则上面加了着重号的一句话应改为"而且 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{A} \times \boldsymbol{B}$ 成一个与原来坐标系手征相同的坐标系". 向量积在物理学中极为重要. 例如设一质点之位置向量与速度向量是 $\boldsymbol{r}$ 和 $\boldsymbol{v}$，则其角速度 $\boldsymbol{\omega} = \boldsymbol{r} \times \boldsymbol{v}$ 就是向量积. 因此，$\boldsymbol{\omega}$ 是赝向量或轴向量. 同样还有角动量 $\boldsymbol{r} \times m\boldsymbol{v}$ 和力矩 $\boldsymbol{r} \times m\boldsymbol{F}$ 都是. 可是更有趣的是转到电磁理论. 我们都知道有电场 $\boldsymbol{E}$ 和磁场 $\boldsymbol{B}$. 上面我们已经看见了，$\boldsymbol{E}$ 是真向量，可是由安培定律，如果有电流经过导线 $\mathrm{d}\boldsymbol{l}$ (电流即运动的电荷)，设电流值为 $i$，则它在 $D$ 点产生的磁场是

$$\mathrm{d}\boldsymbol{B} = i\mathrm{d}\boldsymbol{l} \times \boldsymbol{r}/r^3.$$

(图 7－1－5)由此立刻看到 $\boldsymbol{B}$ 其实是一个赝向量.

不仅如此，我们来看以速度 $\boldsymbol{v}$ 运动着的电荷 $q$ 在电磁场 $\boldsymbol{E}$ 和 $\boldsymbol{B}$ 中所受的力——称为洛伦兹力(Lorentz force)(注意，库伦定律只适用于静电荷)：

$$F = q\left(E + \frac{1}{c}v \times B\right), \qquad c = 光速.$$

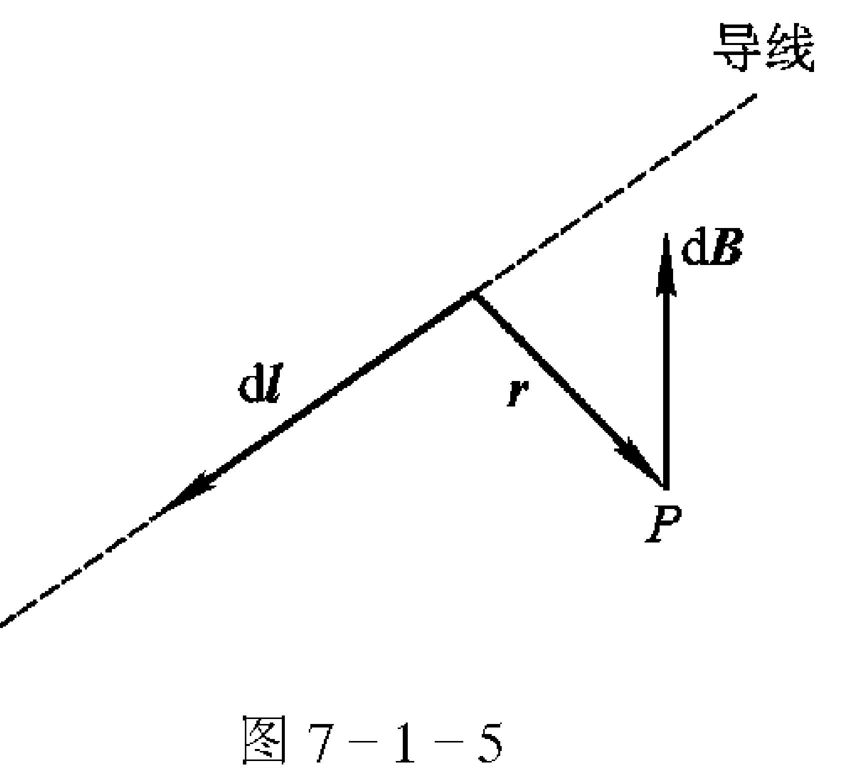



图 7-1-5

它是真向量还是赝向量？第一项当然是真向量.第二项当手征改变时，一方面作为向量积会变一次符号，另一方面 $\boldsymbol{B}$ 作为赝向量又会变一次符号，最终是不变.所以洛伦兹力和牛顿力一样都是真向量.

但是看来电磁场中却存两个性质完全不同的成分：真向量 $\boldsymbol{E}$ 与赝向量 $\boldsymbol{B}$，两个不同性质的东西怎能共处在一起？其实还有一件"怪事"值得在此提一下：细心的读者一定会注意到，我们只讨论三维空间中的向量积，这是为什么？试看上面给的 $\boldsymbol{A}=(a_1,a_2,a_3)$，$\boldsymbol{B}=(b_1,b_2,b_3)$，则 $\boldsymbol{A}\times\boldsymbol{B}$ 的三个分量恰好是 $2\times3$ 矩阵

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \tag{32}$$

的三个二阶子行列式(附以适当的正负号)，于是要问，对于 $n$ 维向量 $\boldsymbol{A}=(a_1,a_2,\cdots,a_n)$，$\boldsymbol{B}=(b_1,b_2,\cdots,b_n)$ 可否取 $2\times n$ 矩阵

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix} \tag{33}$$

的二阶子行列式附以适当符号来构成一个新向量呢？但是现在二阶子行列式的个数是 $\binom{n}{2}=\frac{1}{2}n(n-1)$.如果想构成一个 $n$ 维向量，其充分必要条件是 $\frac{1}{2}n(n-1)=n$，即 $n=3$.所以三维向量的"向量积"$\boldsymbol{A}\times\boldsymbol{B}$ 是一个三维向量，其实是一件"偶然"的事，它一定还另有实质.同样，磁场 $\boldsymbol{B}$ 由安培定律也是一个"向量积"，所以它的"实质"是什么需要深入讨论.到了那时才能对麦克斯韦的电磁理论有一个确切的理解.

向量有真赝之分，标量当然也有.真赝标量之别在于手征变化后变号或不变符号.许多物理量例如质点的质量、光速等等都是真标量.但是例如磁场 $\boldsymbol{B}$ 与位置向量 $\boldsymbol{r}$ 的标量积 $\boldsymbol{r}\cdot\boldsymbol{B}$ 却是一个赝标量.

那么 $\boldsymbol{B}$ 的本质是什么呢？$\boldsymbol{B}$ 和 $\boldsymbol{E}$ 在一起构成一个二阶张量.

**3. 张量**　数学中要研究的量，除了标量，向量(以及标量值，向量值函数)以外，还有许多.把张量也纳入数学研究的对象是一大进步，因为这就大大扩大了数学的视野.人们发现许多在物理学和几何学中极重要的量都是张量.这里我们只从几个例子出发，用有坐标的便于计算的方式引入张量的概念和它们的基本运算规则，其比较系统的不依赖于坐标的讨论则放在§3 中与外形式运算一起来讲.

我们已看到，$n$ 维空间中若赋予了坐标，其向量就是一组 $n$ 个数(本节之中我们限于讨论实的线性空间)$(a_1,\cdots,a_n)$ 并要求它们在坐标变换下按一定规则变化，并视其所遵循的规则称为逆变与协变向量.$k$ 阶张量，有了坐标以后，则是一组 $n^k$ 个数，而在坐标变换下也要求它们按一定规则变化.一个例子是黎曼引进的度量.黎曼的思想是，对一般的非平直的空间，其无限接近的两点之间的距离(这里又涉及无穷小的问题.但读者们不会再有困难)不再是由毕达哥拉斯定理

得出的 $ds^2=dx^2+dy^2$（这与我们在欧氏空间中引入曲线坐标如极坐标、球坐标还不一样.因为现在的空间中根本不一定有 $\mathbf{R}^n$ 中的直角坐标系存在）而是

$$ds^2=g_{ij}(x)dx^i dx^j,\qquad g_{ij}(x)=g_{ji}(x),\tag{34}$$

这里矩阵 $(g_{ij}(x))$ 规定为正定的.

如果我们作一个坐标变换

$$y=y(x),\tag{35}$$

这里 $x(y)$ 是光滑的，而且上式是局部微分同胚，所以雅可比行列式 $\frac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_n)}\neq 0$. 本节开始时我们讲的都是线性变换

$$y=Ax,$$

现在则因所讨论的空间不是 $\mathbf{R}^n$，其中的坐标变换当然不一定是线性变换了.不过在下一节介绍了微分流形及其切空间概念后将会知道，这里没有什么本质区别.我们现在只说(35)式在某定点（不妨设为 $x=0, y(0)=0$）附近的线性部分就是 $y=Ax$，这里 $A=\left.\frac{\partial(y_1,\cdots,y_n)}{\partial(x_1,\cdots,x_n)}\right|_{x=0}$. 知道了这一点以后，下面我们就只讨论变换(35)了.

$ds^2$ 的定义怎样能做到与坐标无关？为了从几何上研究空间的性质而不把坐标系的特点掺进去，这是完全必要的.为此就要求 $(g_{ij}(x))$ 服从一定的要求.这要求就是，若在坐标系 $y$ 之下，$g_{ij}(x)$ 变成了 $\tilde{g}_{kl}(y)$，我们应该有

$$ds^2=\tilde{g}_{kl}(y)dy^k dy^l=g_{ij}(x)dx^i dx^j,$$

以(35)式代入中间一项即得：

$$ds^2=\tilde{g}_{kl}(y)\frac{\partial y^k}{\partial x^i}dx^i\frac{\partial y^l}{\partial x^j}dx^j=\left(\tilde{g}_{kl}(y)\frac{\partial y^k}{\partial x^i}\frac{\partial y^l}{\partial x^j}\right)dx^i dx^j=g_{ij}(x)dx^i dx^j.$$

由 $dx^i$ 的任意性即得

$$\tilde{g}_{kl}(y)\frac{\partial y^k}{\partial x^i}\frac{\partial y^l}{\partial x^j}=g_{ij}(x).\tag{36}$$

注意到 $\frac{\partial x^i}{\partial y^l}\frac{\partial y^k}{\partial x^i}=\delta_l^k$，也可得出

$$\tilde{g}_{kl}(y)=g_{ij}(x)\frac{\partial x^i}{\partial y^k}\frac{\partial x^j}{\partial y^l}.\tag{37}$$

注意，(36)，(37)和前面的(12)，(13)是很相似的，例如(36)中的 $\left(\frac{\partial y^k}{\partial x^i}\right)=A$（注意(36)中是对 $k,l$ 求和的，而(37)中的 $\left(\frac{\partial x^i}{\partial y^k}\right)$ 是 ${}^tA^{-1}$，(37)中是对 $i,j$ 求和的）.

我们再来看一个例子即 $A\times B$. 如果用 $\mathbf{R}^3$ 中的 o.n. 标架 $(\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3)$，注意到从 $\boldsymbol{e}_j$ 到 $\boldsymbol{e}_k$ 所得的平行四边形面积为 $+1$（只要 $i,j,k$ 可由 1,2,3 用轮换排列而得），于是

$$\begin{aligned}\boldsymbol{A}\times\boldsymbol{B}&=a^2b^3\boldsymbol{e}_2\times\boldsymbol{e}_3+a^3b^1\boldsymbol{e}_3\times\boldsymbol{e}_1+a^1b^2\boldsymbol{e}_1\times\boldsymbol{e}_2\\&\quad+a^3b^2\boldsymbol{e}_3\times\boldsymbol{e}_2+a^1b^3\boldsymbol{e}_1\times\boldsymbol{e}_3+a^2b^1\boldsymbol{e}_2\times\boldsymbol{e}_1\\&=(a^2b^3-a^3b^2)\boldsymbol{e}_2\times\boldsymbol{e}_3+(a^3b^1-a^1b^3)\boldsymbol{e}_3\times\boldsymbol{e}_1+(a^1b^2-a^2b^1)\boldsymbol{e}_1\times\boldsymbol{e}_2.\end{aligned}\tag{38}$$

注意到 $\boldsymbol{e}_2\times\boldsymbol{e}_3$ 是 $yz$ 平面上由 $y$ 轴的单位向量转到 $z$ 轴的单位向量所成的平行四边形(其实是正方形)的有符号的面积,则我们可以把(38)式看成一个平面有向面积向三个坐标平面上的分解,也就不妨把这个有向面积看成一种向量的类似物.它有十分类似于平行四边形法则的分解(可惜我们这里采用了 o.n. 标架,而它是与欧几里得度量有关的,平行四边形法则却是与度量无关的.这个问题我们在讲到外形式时还要讲),最好是令例如 $a^jb^k-a^kb^j=2T^{jk}$,则(38)可写为

$$\boldsymbol{A}\times\boldsymbol{B}=T^{jk}\boldsymbol{e}_j\times\boldsymbol{e}_k, \tag{39}$$

这里 $j,k$ 可以由 1 到 3 任意取值,而不必如(38)那样一定要求 $(i,j,k)$ 成为轮换对称.这样自动地有

$$T^{jk}=-T^{kj},\qquad T^{ii}=0. \tag{40}$$

$T^{jk}$ 称为 $\boldsymbol{A}\times\boldsymbol{B}$ 的分量.

现在作一个坐标变换 $y=Ax$,使原来的基底成为新基底 $\boldsymbol{f}_i=\Lambda\boldsymbol{e}_i$,于是 $\boldsymbol{e}_i=\lambda_i^j\boldsymbol{f}_j$,而 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别成为

$$\boldsymbol{A}=a^i\boldsymbol{e}_i=a^i\lambda_i^j\boldsymbol{f}_j=\alpha^j\boldsymbol{f}_j,\qquad \alpha^j=\lambda_i^ja^i,$$
$$\boldsymbol{B}=b^i\boldsymbol{e}_i=b^i\lambda_i^j\boldsymbol{f}_j=\beta^j\boldsymbol{f}_j,\qquad \beta^j=\lambda_i^jb^i,$$

于是在新坐标系下

$$\boldsymbol{A}\times\boldsymbol{B}=(\alpha^2\beta^3-\alpha^3\beta^2)\boldsymbol{f}_2\times\boldsymbol{f}_3+(\alpha^3\beta^1-\alpha^1\beta^3)\boldsymbol{f}_3\times\boldsymbol{f}_1+(\alpha^1\beta^2-\alpha^2\beta^1)\boldsymbol{f}_1\times\boldsymbol{f}_2.$$

现在新的分量是例如 $\alpha^2\beta^3$ 等等,而有 $\alpha^2\beta^3=\lambda_i^2\lambda_j^3a^ib^j$,等等.这个式子显然与逆变向量坐标的变换公式很相近.$\{g_{ij}(x)\}$和 $\boldsymbol{A}\times\boldsymbol{B}$ 都是我们要讲的张量.它们是二阶张量,因为它们的分量或坐标都依赖于两个指标.前者是协变张量,后者是逆变张量,其理由自明.前者是对称张量,因为 $g_{ij}(x)=g_{ji}(y)$,后者是反对称张量.因为(40)告诉我们 $T^{kj}=-T^{jk}$.

按这里所讲的,我们可以给出一般的 $k$ 阶张量的定义,我们先令 $k=p+q$,$p,q$ 为非负整数.我们要构造出一个 $p$ 次逆变,$q$ 次协变的张量.这种张量称为 $(p,q)$ 型张量.于是有

**定义 2**　设有 $n^k$ 个光滑函数 $T_{j_1,\cdots,j_q}^{i_1,\cdots,i_p}(x)$,这里 $i_1,\cdots,i_p$ 与 $j_1,\cdots,j_q$ 均互相独立地从 1 到 $n$ 取值.若作一个局部微分同胚(35),使得在新坐标系下有 $n^{p+q}$ 个函数 $\tilde{T}_{l_1,\cdots,l_q}^{k_1,\cdots,k_p}(y)$,而与原来的函数组 $T_{j_1,\cdots,j_q}^{i_1,\cdots,i_p}(x)$ 之间有以下关系式

$$\tilde{T}_{l_1,\cdots,l_q}^{k_1,\cdots,k_p}(y)=\frac{\partial y^{k_1}}{\partial x^{i_1}}\cdots\frac{\partial y^{k_p}}{\partial x^{i_p}}\frac{\partial x^{j_1}}{\partial y^{l_1}}\cdots\frac{\partial x^{j_p}}{\partial y^{l_p}}T_{j_1,\cdots,j_q}^{i_1,\cdots,i_p}(x), \tag{41}$$

就说 $\{T_{j_1,\cdots,j_q}^{i_1,\cdots,i_p}(x)\}$ 是一个 $k$ 阶 $(p,q)$ 型张量,更确切地说是一个 $p$ 阶逆变,$q$ 阶协变张量,$i_1,\cdots,i_p$ 和 $j_1,\cdots,j_q$ 分别称为逆变和协变指标.

很清楚,向量就是一阶张量.逆变与协变向量分别是(1,0)型与(0,1)型张量.标量则是 0 阶张量,或者说是(0,0)型张量.

有一个特别重要的(1,1)型张量 $\delta_j^i$,称为克罗内克(Kronecker)张量,它的特点在于其不变性:若它在 $y$ 坐标下的分量是 $\tilde{\delta}_l^k$,则

$$\tilde{\delta}_l^k=\delta_l^k.$$

事实上

$$\tilde{\delta}_l^k = \frac{\partial y^k}{\partial x^i}\frac{\partial x^j}{\partial y^l}\delta_j^i = \frac{\partial y^k}{\partial x^i}\frac{\partial x^i}{\partial y^l} = \delta_l^k.$$

我们说它是一个不变张量.另一个重要的张量是 $n$ 维空间中的 $(0,n)$ 型张量——列维－齐维塔(Levi－Civita)张量,它在 $\mathbf{R}^n$ 的一个正交坐标系下的分量是

$$\varepsilon_{i_1,\cdots,i_n} = \begin{cases} 1, & 若\begin{pmatrix} 1,\cdots,n \\ i_1,\cdots,i_n \end{pmatrix}是偶置换, \\ -1, & 若\begin{pmatrix} 1,\cdots,n \\ i_1,\cdots,i_n \end{pmatrix}是奇置换, \\ 0, & 若\ i_1,\cdots,i_n\ 中有相等者. \end{cases} \tag{42}$$

它只在正交变换下有不变性:考虑 $\mathbf{R}^n$ 上的一个旋转变换 $y=Ax$,这里 $A\in SO(n)$.令

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

则$\dfrac{\partial y^i}{\partial x^j}=a_{ij}$,而且 $\det A=1$,另一方面若记 $\varepsilon_{i_1,\cdots,i_n}$ 在 $y$ 坐标系下的分量为$\tilde{\varepsilon}_{j_1,\cdots,j_n}$,则

$$\tilde{\varepsilon}_{j_1,\cdots,j_n} = \sum_i \varepsilon_{i_1,\cdots,i_n}\frac{\partial y^{j_1}}{\partial x^{i_1}}\cdots\frac{\partial y^{j_n}}{\partial x^{i_n}}.$$

这里我们没有区分协变与逆变坐标,因为在正交变换下二者没有区别.

由行列式的定义

$$\sum_i \varepsilon_{i_1,\cdots,i_n} a_{j_1,i_1}\cdots a_{j_n,i_n} = \begin{vmatrix} a_{j_1 1} & \cdots & a_{j_1 n} \\ \\ a_{j_n 1} & \cdots & a_{j_n n} \end{vmatrix}.$$

所以若 $\begin{pmatrix} 1,\cdots,n \\ j_1,\cdots,j_n \end{pmatrix}$ 是偶置换,此行列式等于 $\det A=1$;若 $\begin{pmatrix} 1,\cdots,n \\ j_1,\cdots,j_n \end{pmatrix}$ 为奇置换,此行列式等于 $-\det A=-1$;若 $j_1,\cdots,j_n$ 中有二者相等,此行列式有两行相同,因而为 0.

在一般坐标变换下,列维－齐维塔张量没有不变性.

下面讨论张量的代数运算.为了书写方便,我们不写出关于一般 $(p,q)$ 型张量的公式而只写出对某特定类型张量的相应公式.一般公式自然容易得出.首先是加法,设有两个同型的张量 $A_{j_1}^{i_1 i_2}$ 与 $B_{j_1}^{i_1 i_2}$,则二者可以求和,其和也是同型张量,而其分量是

$$C_{j_1}^{i_1 i_2} = A_{j_1}^{i_1 i_2} + B_{j_1}^{i_1 i_2}. \tag{43}$$

其次是张量积,仍设有两个张量,$A_{j_1}^{i_1 i_2}$ 与 $B_{j_2 j_3}^{i_3}$,分别是 $(2,1)$ 型与 $(1,2)$ 型的,则定义其张量积 $C=A\otimes B$ 为一个 $(2+1,1+2)$ 型张量

$$C_{j_1 j_2 j_3}^{i_1 i_2 i_3} = A_{j_1}^{i_1 i_2}\cdot B_{j_2 j_3}^{i_3}. \tag{44}$$

有时我们又会把它写成 $A_{j_1}^{i_1 i_2} \otimes B_{j_2 j_3}^{i_3}$. 应该注意，一般说来 $A \otimes B \neq B \otimes A$. 即以(44)为例，如果把(44)中的 $C$ 理解为 $A \otimes B$，则例如有

$$C_{1\,2\,3}^{1\,2\,3} = A_1^{1\,2} \cdot B_{2\,3}^3,$$

但若将 $C$ 理解为 $B \otimes A$，则相应有

$$C_{1\,2\,3}^{1\,2\,3} = B_{1\,2}^1 \cdot A_3^{2\,3}.$$

二者确实是不相同的.

张量积有一个特例，即标量 $\lambda$ 可以看成0阶张量. 这样对任一张量，$A$ (其分量为 $A_{j_1}^{i_1 i_2}$)就可以定义 $\lambda \otimes A$ (其分量为 $\lambda A_{j_1}^{i_1 i_2}$)仍为同型张量. 由上所述，同型的张量对于例如实数域中的 $\lambda$，形成一个线性空间. 我们记之为 $T_{j_1}^{i_1 i_2}$，不同型的张量对于直和形成一个非交换的代数，并以 $\otimes$ 为其中的乘法，这个代数称为张量代数 $T$. 关于张量代数与张量空间的代数结构，我们将在第三节去讨论.

但是有两个常用的记号应该在此提到. 我们知道，一个向量(不论是协变的还是逆变的)都可以看成一个一阶张量. 例如，$\boldsymbol{A} = \sum_{i=1}^{n} a_i \boldsymbol{e}_i$，$\boldsymbol{B} = \sum_{j=1}^{n} b_j \boldsymbol{e}_j$. 求二者的张量积，则可得到一个二阶张量.

$$\boldsymbol{A} \otimes \boldsymbol{B} = \sum_{i,j=1}^{n} a_i b_j \boldsymbol{e}_i \otimes \boldsymbol{e}_j.$$

如果我们把 $\boldsymbol{e}_i \otimes \boldsymbol{e}_j$ 看成这个二阶张量空间的基底(后面我们也将证明它)，则 $\boldsymbol{A} \otimes \boldsymbol{B}$ 可以看成与 $\{a_i b_j\}$ 一一对应，或者就与下面的 $n \times n$ 矩阵一一对应. 因此，我们就简单地写作

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_1 b_1 & \cdots & a_1 b_n \\ & & \\ a_n b_1 & \cdots & a_n b_n \end{pmatrix}. \tag{45}$$

前面我们讲到黎曼的度量张量时，也曾把它写成以下几种形式之一

$$\mathrm{d}s^2 = g_{ij}(x)\mathrm{d}x^i \mathrm{d}x^j,$$
$$\{g_{ij}(x)\},$$
$$(g_{ij}(x)),$$

甚至可以写成

$$\mathrm{d}s^2 = g_{ij}(x)\mathrm{d}x^i \otimes \mathrm{d}x^j.$$

这纯粹是一个记号问题，不过它在物理学和力学中常会见到这种情况，并且把(45)这样的二阶张量称为并矢(dyad).

另一点应该提到的是在泛函分析特别在广义函数论中，张量积这个词又有另一个用法. 如果 $u(x)$ 与 $v(x)$ 都是定义在 $\Omega$ 中的函数，我们定义其张量积为 $(u \otimes v)(x_1, x_2) = u(x_1)v(x_2)$. 这样做的理由在于，$A$ 与 $B$ 本来属于同一个线性空间 $E$，其基底本来是同样的 $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$. 但在考虑张量积时，我们认为 $A$ 和 $B$ 各在自己的 $E$ 中，而这两个 $E$ 是互相独立的. 因此，我们构造了另一个线性空间 $E \otimes E$，它是 $n^2$ 维的，而以 $\{\boldsymbol{e}_i \times \boldsymbol{e}_j\}$，$i, j = 1, \cdots, n$ 为基底，把这个思想用于函数，$u(x)$ 与 $v(x)$ 可能属于同一空间，例如 $C_0^\infty(\Omega)$，但我们要求二者独立变化互不相涉. 这样做最

简单的方法自然是对 $u$ 和 $v$ 的自变量采用不同的记法，分别把它们写成 $u(x_1)$ 和 $v(x_2)$，并把通常的求积 $u(x_1)v(x_2)$ 写成张量积 $(u\otimes v)(x_1,x_2)$. 这样做，在代数结构上会得到一些好处，我们就不再去讲了. 同样，在量子力学中，两个函数之并矢，亦即 $u\otimes v$，有时写成

$$(u\otimes v)(x_1,x_2)=|u><v|.$$

这是狄拉克(Dirac)的记号，它时常是很方便的.

张量所特有的，也是十分重要的代数运算是它的"缩并"(contraction). 它的意思是说在一个张量积，甚至在一个张量(视为该张量与 1 的张量积)中，若令一个协变指标与另一个逆变指标相等并对之求和，则将由原来的 $(p,q)$ 型张量得出一个 $(p-1,q-1)$ 型张量. 因此，该张量的阶将下降 2 阶. 例如在 $A^{ij}B_k$ 中令 $j=k$ 而且对 $j$ 相加，用求和规定将得到

$$A^{ij}B_j=C^i,$$

则 $C^i$ 是一个 (1,0) 型张量，这个证明是很容易的. 因为若记这些量在 $y$ 坐标系下的表示为 $\tilde{A}^{ij}$，$\tilde{B}_k$ 和 $\tilde{C}^i$，则

$$\tilde{C}^i=\tilde{A}^{ij}\tilde{B}_j=\frac{\partial y^i}{\partial x^k}\frac{\partial y^j}{\partial x^l}\frac{\partial x^m}{\partial y^j}A^{kl}B_m$$

$$=\frac{\partial y^i}{\partial x^k}\delta_l^m A^{kl}B_m=\frac{\partial y^i}{\partial x^k}C^k.$$

所以 $C^i$ 是一个 (1,0) 型张量.

其实缩并是我们已经见到过的. 例如一个 (1,1) 型张量 $A_j^i$，前面我们已说过，可以表示为一个矩阵

$$A=\begin{pmatrix}a_1^1 & \cdots & a_n^1\\ & & \\ a_1^n & \cdots & a_n^n\end{pmatrix},$$

则其缩并就是

$$a_i^i=a_1^1+\cdots+a_n^n.$$

它恰好就是矩阵 $A$ 之迹(trace) tr $A$：

$$\mathrm{tr}\,A=a_i^i.$$

缩并的另一个表现就是内积. 一个逆变向量 $\mathbf{A}=(A^1,\cdots,A^n)$ 与一个协变向量 $\boldsymbol{B}=(B_1,\cdots,B_n)$ (视向量为一阶张量)之内积，其实就是对 $\mathbf{A}\otimes\boldsymbol{B}$ 作缩并的结果：

$$\mathbf{A}\cdot\boldsymbol{B}=\sum_j(A^iB_j)\mid_{j=i}.$$

利用缩并还可以得到一个决定某一组量是否张量的方法. 设有 $n^2$ 个量，它们在 $x$ 坐标下的表示为 $A_{ik}$，若对此坐标系下的任意逆变向量 $X^k$，均可证明 $A_{ik}X^k$ 是一个 (0,1) 型张量，则 $A_{ik}$ 必为一个 (0,2) 型张量，实际上若换到 $y$ 坐标系而 $A_{ik}$ 与 $X^k$ 分别成了 $\tilde{A}_{ik}$ 与 $\tilde{X}^k$，则

$$\tilde{A}_{ik}\tilde{X}^k=\frac{\partial x^l}{\partial y^i}A_{lm}X^m.$$

但是

$$\tilde{X}^k=\frac{\partial y^k}{\partial x^m}X^m,$$

代入上式有

$$\left(\tilde{A}_{ik}\frac{\partial y^k}{\partial x^m}-\frac{\partial x^l}{\partial y^i}A_{lm}\right)X^m=0.$$

因为此式对任意逆变向量 $X^m$ 均成立,故有

$$\tilde{A}_{ik}\frac{\partial y^k}{\partial x^m}=A_{lm}\frac{\partial x^l}{\partial y^i}.$$

双方乘以$\frac{\partial y^i}{\partial x^p}$并对 $i$ 求和,利用$\frac{\partial x^l}{\partial y^i}\frac{\partial y^i}{\partial x^p}=\delta^l_p$,有

$$\tilde{A}_{ik}\frac{\partial y^k}{\partial x^m}\frac{\partial y^i}{\partial x^p}=A_{lm}\delta^l_p,$$

亦即

$$\tilde{A}_{ik}\frac{\partial y^k}{\partial x^m}\frac{\partial y^i}{\partial x^p}=A_{pm}.$$

所以 $A_{lk}$ 是一个(0,2)型张量.

作为本节之结束,我们来回答为什么坐标变换不写成(6)而要写成 $x^j=\lambda^j_iy^i$? 这不但为了应用爱因斯坦求和规定,而且要保证在变换前后有相同的协变与逆变性质.

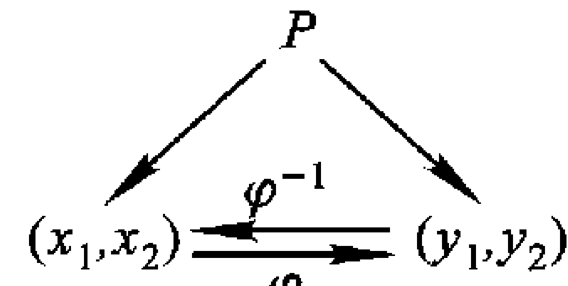
## §2　微分流形

**1. 基本定义和例子**　现在我们开始比较系统地讨论弯曲的空间,并在弯曲的空间上建立起微积分学.

当然,我们需要把研究的对象明确起来.一方面要求足够广泛,能包括我们可能会遇到的对象,主要是在物理学与几何学中的对象;另一方面又不能没有界限以至于会因为不能不处理一些"病态"的对象,而把我们已经掌握得很彻底的微积分理论丢开而另起炉灶.这两个方面分寸如何掌握最终是由物理学和几何学的需要决定的.科学的发展证明,一个比较适当的界限可以划到微分流形为止.

什么是一个微分流形? 我们从一个实例开始.绘制地图时,我们没有办法把地球球面展开成一个平面而且保持每一个国家、地区"形状不变".于是我们把地球分成许多小的区域,把各个区域都投影到某一个平面区域上,而且各个不同区域采用不同的投影方法,例如麦卡托(G. Mercator)(不是第二章讲到对数函数时提到的那个麦卡托)在1569年提出的麦卡托投影法.但最易懂的是柱面投影法:把地球看作一个球面而沿赤道用一个圆柱包围起来,并从球心将球面上各点投影到柱面上,再把柱面打开成一个平面就得到一张世界全图(世界全图称为 atlas,此词来自希腊神话,是麦卡托首先用这个词的.分国分省地图都不能称为 atlas.讲一下这个故事对理解下文有好处).这个投影法的好处是子午线和纬圈变成了互相垂直的直线族,而且在赤道附近,区域的"形状"保持得很好,但是例如北极被投影到无穷远处,而且高纬度的国家如俄罗斯、加拿大都大得出奇,很不"真实"了.另一种投影法例如球极投影(前面也讲过),如果从南极开始把球面向北极的切平面投影,则子午线变成了过北极的直线,纬圈仍是圆周,很像极坐标,它的好处是:在北极附近,图形很"真实",而且经过北极的大圆(即子午线)都是直线,角度大小也不变,但是南极被

映到无穷远处去了.因此地图学的实践是在不同区域采用不同的投影,而在采用了不同投影的两个区域之公共部分,则要研究其“迁移”(transition,这是一个数学名词而不是制图学名词).事实上如果在第一种投影之下平面上的点用笛卡儿坐标$(x_1,x_2)$表示,第二种投影之下得到$(y_1,y_2)$,则同一点$P$就既有$(x_1,x_2)$又有$(y_1,y_2)$,用箭头图示如下:




从这个图看到,在$(x_1,x_2)$与$(y_1,y_2)$之间有一个映射$\varphi$(其逆映射为$\varphi^{-1}$),而用坐标表示成为

$$y_1=\varphi_1(x_1,x_2),y_2=\varphi_2(x_1,x_2). \tag{1}$$

这一对函数就称为“迁移函数”(transition functions,这又是数学名词而不是制图学的名词).如果我们想在地球上做微积分,当然应该要求它们是光滑函数(但远不止于此).从这里我们汲取了以下的研究弯曲空间的基本思想:

1. 把研究对象分成许多小的部分,而每一部分都可以投影到欧氏空间$\mathbf{R}^n$内,因而可以进行局部的讨论.

2. 在两个部分相交的地方,通过研究迁移函数来实现从一个部分到另一个部分的转换.

现在就可以介绍微分流形的概念了.但是立刻就遇到一个问题,在讨论地球表面的各种投影以及在讨论更一般的曲面时,我们的对象是很清楚的:它是$\mathbf{R}^3$的曲面,它的弯曲是相对于包含它的空间——所谓包含空间(ambient space)——而言的.但是本章一开始就指出,黎曼的基本思想是:空间本身就是弯曲的,用不着有包含空间作衬托.而其弯曲就表现在,其各点附近就距离而言并非均匀的,因而度量$\mathrm{d}s^2=g_{ij}(x)\mathrm{d}x^i\mathrm{d}x^j$随$x$而异.现在我们也不能设想还有什么包含空间,也不知道有没有度量.因而我们一开始只有把研究对象$M$规定是一个拓扑空间.

然后,我们把$M$划分成许多小片,具体说,设有若干个开集——注意,拓扑空间中是可以定义开集$\{U_\alpha\}$的,使$M=\bigcup\limits_\alpha U_\alpha$.对于每一个$U_\alpha$,我们都假设有一个“投影”$\varphi_\alpha:U_\alpha\to\mathbf{R}^n$,从数学上说,$\varphi_\alpha$是一个同胚,因此$\varphi_\alpha(U_\alpha)\subset\mathbf{R}^n$是一个开集,甚至是一个区域.$U_\alpha$称为一个区图(chart,这是一个古老的英文字,下面的解释现在已不用了:一片水域或陆地的地图),而其集$\{U_\alpha\}$称为$M$的一个图册(atlas),这都是借用了制图学中的名称.$\varphi_\alpha(U_\alpha)\subset\mathbf{R}^n$,$\mathbf{R}^n$中是有坐标的,我们就借用这个坐标作为$U_\alpha\subset M$中点的坐标.这种坐标称为局部坐标,$\varphi_\alpha$称为坐标映射(有时也就称$\varphi_\alpha$为局部坐标).我们不妨设$U_\alpha$是连通的,因若不然,我们可以把$U_\alpha$的各个连通分支各当作一个$U_\alpha$.因为$\varphi_\alpha$是同胚,所以$\varphi_\alpha(U_\alpha)\subset\mathbf{R}^n$也是连通开集——所谓区域就是连通开集.这样我们看到,通过同胚$\varphi_\alpha$可以把$\mathbf{R}^n$的拓扑性质——连通性、局部紧性——都移植到$M$上.结果,所有我们将要定义的微分流形$M$都是局部连通与局部紧的.可见,微分流形将不会是一个过于空泛的概念.

如果$U_1\cap U_2\neq\varnothing$,则不论是否$\varphi_1(U_1)\cap\varphi_2(U_2)\neq\varnothing$,在$\varphi_1(U_1)$和$\varphi_2(U_2)$的局部坐标$(x_1,\cdots,x_n)$与$(y_1,\cdots,y_n)$之间都会得到迁移函数作为一个映射:

$$\varphi_2\circ\varphi_1^{-1}:\varphi_1(U_1)\to\varphi_2(U_2).$$

用局部坐标来表示就是

$$\begin{aligned} y_i &= (\varphi_2 \circ \varphi_1^{-1})(x_1, \cdots, x_n) \\ &= \psi_i(x_1, \cdots, x_n). \end{aligned} \tag{2}$$

为了使得在 $M$ 上有微分学，就需要一个至关重要的相容性条件：(2)是微分同胚.重要的是，我们不妨取 $\varphi_1(U_1)$，$\varphi_2(U_2)$充分小，使得这个微分同胚条件可以用坐标来表示为：$\psi \in C^\infty$，$\psi^{-1} \in C^\infty$，且

$$\det \frac{\partial(y_1, \cdots, y_n)}{\partial(x_1, \cdots, x_n)}\bigg|_{\varphi_1(U_1)} \neq 0. \tag{3}$$

"充分小"这样的限制词是必要的，如若不然，仅由雅可比行列式(3)不为 0 尚不足以保证 $\varphi_2 \circ \varphi_1^{-1} = \psi$ 为一微分同胚.

最后我们还要加上两个条件：首先 $M$ 是豪斯多夫空间，这不仅使我们避免了与许多"病态"的对象打交道，而且有了它，再加上第二个条件即 $M$ 适合"第二可数性公理"，即存在可数多个开集$\{U_i\}$ $i=1,2,\cdots$，使 $M$ 中任一开集必是若干个(可能是可数多个)$U_i$ 之并，就可以构造出许多极为有用的重要的对象.首先是单位分解(见第三章)，有了它，才能把有关 $M$ 的问题例如积分问题归结为在各个区图 $U_\alpha$ 上的相应问题.至今我们没有说，$M$ 上是否可以按照黎曼的要求构造出一个度量 $ds^2 = g_{ij}(x)dx^i dx^j$，而在有了单位分解以后，就可以证明 $M$ 上确可定义黎曼的度量.证明见§5.这以后 $M$ 之每个连通分支均各将成为一个度量空间，这一点我们则不加证明了.

有了这些讨论，我们就可以正式地给出微分流形的定义.

**定义 1** 设 $M$ 为一适合第二可数性公理的豪斯多夫拓扑空间，而且有由连通开集$\{U_\alpha\}$所成的覆盖，使得对每个 $U_\alpha$ 存在同胚映射 $\varphi_\alpha: U_\alpha \to \mathbf{R}^n$，而且当 $U_\alpha \cap U_\beta \neq \varnothing$ 时迁移函数 $\varphi_\beta \circ \varphi_\alpha^{-1}$：$\varphi_\alpha(U_\alpha) \to \varphi_\beta(U_\beta)$为一($C^\infty$)微分同胚.这时称 $M$ 为一 $n$ 维($C^\infty$)微分流形，$U_\alpha$ 称为一个区图(或坐标邻域)，$\{U_\alpha\}$称为图册，$\varphi_\alpha(U_\alpha) \subset \mathbf{R}^n$ 在 $\mathbf{R}^n$ 中的坐标称为 $M$ 中 $U_\alpha$ 之局部坐标，$\varphi_\alpha$ 称为坐标映射，$n$ 称为 $M$ 之维数.

**注 1** 定义中只要求$\{U_\alpha\}$能覆盖 $M$，而有时需要把一切适合 $\varphi(V) \subset \mathbf{R}^n$ 的同胚 $\varphi$ 与连通开集 $V$ 也都列入$\{\varphi_\alpha\}$和$\{U_\alpha\}$之中，但要 $\varphi \circ \varphi_\alpha^{-1}$ 为微分同胚，以得到一个极大图册(maximal atlas).极大图册的存在性是需要而且可以证明的.这种极大图册称为一个微分构造.于是产生一个问题：同一个 $M$ 上是否有唯一的微分构造.20 世纪 60 年代，米尔诺(J. Milnor)证明了 $S^7$ 上就有互不相容的微分构造存在，即一个构造中的 $\varphi_\alpha$ 与另一构造中的 $\varphi_\beta$ 能使 $\varphi_\alpha \circ \varphi_\beta^{-1}$ 不是微分同胚.也有人证明了确有没有微分构造的流形存在.20 世纪 80 年代，英国数学家唐纳逊(S. K. Donaldson)证明了 $\mathbf{R}^4$ 中即有互不相容的微分构造，这件事与理论物理有密切的关系.

**注 2** 上面我们讲的是 $C^\infty$ 微分流形，即指迁移函数是 $C^\infty$ 函数.这是最常见的情况，一般数学文献中凡讲到光滑，总是指的 $C^\infty$，所以也时常略去 $C^\infty$(甚至光滑)字样.但实际上还有迁移函数只是 $C^k$ 函数的情况.这时流形就称为 $C^k$ 流形，特别是 $C^0$ 流形，其实并没有可微性.它的性质与其它 $C^k$ 流形可以很不相同.例如可以问，每个微分构造是否只有一个维数？确实如此，但对于 $C^k$ 流形，当 $k>0$ 时证明并不难，$k=0$ 时也是可以证明的，但要证明维数不变需证明著名的区域不变性定理：

**定理 1** 若 $n \neq m$，则 $\mathbf{R}^n$ 之一开集不能同胚于 $\mathbf{R}^m$ 的开集.

这个著名定理是布劳威尔证明的，要用到很细致的同调理论工具，但在 $k>0$ 时却不难证明.由此，我们至少看到，维数是刻画一个流形的整体不变量；对一个微分流形找出尽可能多的整体不变量是很重要的事.这类问题我们就说到这里为止.

有时，迁移函数是实解析函数(一个函数在某点附近是实解析的，即在此点附近可以展开为收敛的幂级数)，实解析函数类时常记为 $C^{\omega}$，实解析流形也就记为 $C^{\omega}$ 流形.读者会问，如果把上面讲的 $n$ 维实空间 $\mathbf{R}^n$ 换成 $n$ 维复空间，而迁移函数从可微函数换成全纯函数，是否可以得到相应的复流形理论？是，而且复流形与实流形很不相同，我们完全不能涉及.

$M$ 上有了微分结构以后，就可以在其上建立种种有关微分学的概念.这里一定要注意，所有这些概念都必须是与局部坐标之选取无关的.我们先确定一下什么是 $M$ 上的可微函数 $f$.如果 $f$ 定义在 $M$ 上某点 $P$ 之邻域中，不失一般性，可以设此邻域即为一坐标邻域 $U$.于是 $f\circ\varphi^{-1}$ 即成为 $\varphi(U)\subset\mathbf{R}^n$ 上的函数.如果 $f\circ\varphi^{-1}$ 是 $C^{\infty}$ 可微的，就说 $f$ 在 $P$ 附近是 $C^{\infty}$ 可微的.这个定义显然与 $U$ 和 $\varphi$ 的选择无关，因为若选 $P$ 的另一个邻域 $V$ 与相应的局部坐标 $\psi$，但在 $U\cap V$ 上 $\psi\circ\varphi^{-1}:\varphi(U\cap V)\to\psi(U\cap V)\subset\psi(U)$ 是一个微分同胚，因而是 $C^{\infty}$ 的，$f\circ\varphi^{-1}$ 也是 $C^{\infty}$ 函数，所以 $f\circ\psi^{-1}=(f\circ\varphi^{-1})\circ(\varphi\circ\psi^{-1})$ 至少在 $P$ 的一个较小的邻域 $U\cap V$ 上是 $(C^{\infty})$ 可微的.

下面我们来举一些 $C^{\infty}$ 微分流形的例子.

**例 1** $M=\mathbf{R}^n$.这是平凡的，因为现在的图册 $\{U_\alpha\}$ 只含一个元 $U=\mathbf{R}^n$ 即可，而坐标映射就是恒等映射：$\varphi_\alpha=\mathrm{id}$.我们这里没有去讨论什么是极大图册——微分构造.讨论那个问题对我们目前并无好处.

其次令 $M$ 为 $\mathbf{R}^n$ 中任一开集 $U$.这时仍有 $\{U_\alpha\}=\{U\}$，$\varphi_\alpha=\mathrm{id}|_U$，我们说 $U$ 是 $\mathbf{R}^n$ 的开子流形.这个概念与下面要讲到的闭子流形是不一样的，开子流形的维数仍为 $n$.

读者会问，半空间 $H^n=\{x\in\mathbf{R}^n,x_n\geqslant 0\}$ 是不是微分流形？一方面我们很需要把它纳入我们的视野，因为例如在积分理论中考虑斯托克斯(Stokes)定理(格林公式是它的特例)时，必须要讨论 $H^n$ 上的积分.另一方面它又不是流形，因为它有一个"边缘" $x_n=0$，其上的点没有任何可以同胚于 $\mathbf{R}^n$ 的邻域.因此，我们要以 $H^n$ 为模型来扩大微分流形的定义.

**定义 2** 设 $M^+$ 是一适合第二可数性公理的豪斯多夫拓扑空间.$\{U_\alpha\}$ 是覆盖 $M^+$ 的开集族，$\bigcup\limits_\alpha U_\alpha=M^+$，而对每个 $U_\alpha$ 均有同胚 $\varphi_\alpha:U_\alpha\to\mathbf{R}^n$ 或 $H^n$，使得当 $U_\alpha\cap U_\beta\neq\varnothing$ 时，$\varphi_\beta\circ\varphi_\alpha^{-1}:\varphi_\alpha(U_\alpha)\to\varphi_\beta(U_\beta)$ 为一微分同胚.这时称 $M^+$ 为带边的微分流形.$\{x\in H^n,x_n=0\}$ 在 $M^+$ 中的原像称为 $M^+$ 的边缘，记作 $\partial M^+$.

**注** 一个带边流形 $M^+$ 之边缘与一个拓扑空间的子空间的边界点是不同的概念.例如，$\mathbf{R}^2$ 中的去了圆心的闭圆盘 $0<r\leqslant 1$ 的边缘只是单位圆周 $r=1$，但其边界点之集合还多了一个点 $r=0$——圆心.

注意，这个定义中 $M^+$ 就是拓扑空间的全空间，因此不存在 $M^+$ 以外的点的问题.也因此，$U_\alpha$ 作为 $M^+$ 中的开集允许含有 $\partial M^+$ 上之点.但是 $H^n$ 又是 $\mathbf{R}^n$ 的子空间，并赋有子空间拓扑，因此 $\{x_n=0\}$ 上之点虽然就 $H^n$ 是 $\mathbf{R}^n$ 之子集而言是 $H^n$ 之边界点，按子空间拓扑而言则可以含于 $H^n$ 之开子集 $\varphi_\alpha(U_\alpha)$ 中.由同样的理由，$\varphi_\beta\circ\varphi_\alpha^{-1}$ 可以一直光滑到 $x_n=0$ 处，我们通常就说，$\varphi_\beta\circ\varphi_\alpha^{-1}$ "光滑到边".现在我们问，把 $H^n$ 作为 $\mathbf{R}^n$ 之子集来看，可否找到"更大"的子集 $K$，包含 $H^n$ 而

且跨越到 $x_n<0$ 处，使 $\varphi_\beta\circ\varphi_\alpha^{-1}$ 可以光滑地拓展到 $K$ 上？一般说来，设 $f(x)$ 是闭集 $H$ 上的光滑函数，$K\supset H$ 是一开集，能否找到一个定义在 $K$ 上的光滑函数 $F(x)$，使在 $H$ 上 $F(x)=f(x)$？$F(x)$ 称为 $f(x)$ 由 $H$ 到 $K$ 上的光滑延拓．是否可能延拓与 $\partial H$ 的构造有密切的关系．作为带边流形的模型 $H^n$ 之边缘构造特别简单，就只是一个超平面 $x_n=0$，比较一般的情况，则可以考虑 $M^+$ 局部也可以用 $f(x)\geqslant 0$ 来刻画，这里 $f(x)\in C^\infty$，而且 $\mathbf{grad}\ f\neq 0$．且 $M^+$ 局部地位于 $\partial M^+=\{x; f(x)=0\}$ 之一侧．在这个情况下有著名的 Seeley 延拓定理：

**Seeley 延拓定理**　设 $f(x)\in D(H^n)$，则必有一个连续映射 $\rho: D(H^n)\to D(\mathbf{R}^n)$，使

$$(\rho f)(x)=f(x),\qquad x\in H^n.$$

由于有这个定理，上述迁移函数，$\varphi_\beta\circ\varphi_\alpha^{-1}$ 可以在 $x_n<0$ 的某一区域中也光滑，且是微分同胚，这对我们讨论带边流形有很大的方便．

这个定理的证明我们就略去了，只是要提醒一下，如果 $\partial M^+$ 不是如此简单，哪怕 $M^+$ 例如是一个象限，角点的出现也会使函数的延拓困难不少．

**例 2　无奇点的光滑超曲面**　设有 $\mathbf{R}^n$ 中一个光滑的超曲面，它由一个方程

$$f(x_1,\cdots,x_n)=0 \tag{4}$$

定义，这里 $f$ 是其自变量在 $\mathbf{R}^n$ 之某区域 $\Omega$ 内的 $C^\infty$ 函数，我们设它是无奇点的，即在 $\Omega$ 上

$$\mathbf{grad}\ f(x)\neq 0,$$

因此在每一点至少有一个偏导数不为 0．设在某个开集 $U$ 中，$\dfrac{\partial f}{\partial x_n}\neq 0$ 而可以应用隐函数定理得出

$$x_n=\varphi(x_1,\cdots,x_{n-1}),\qquad (x_1,\cdots,x_{n-1})\in\omega.$$

这里，适合以上方程的 $(x_1,\cdots,x_{n-1},x_n)\in U$，从而所有这种 $U$ 构成超曲面 (4) 的开覆盖，而且有坐标映射

$$\Phi:\{(x_1,\cdots,x_{n-1},\varphi(x_1,\cdots,x_{n-1}))\in U,(x_1,\cdots,x_{n-1})\in\omega\}\to\omega\subset\mathbf{R}^{n-1}.$$

于是我们需要证明的只是，例如有两个这样的 $U$ 和 $V$，在 $V$ 中 $\dfrac{\partial f}{\partial x_{n-1}}\neq 0$，且有相应的坐标映射：$\Psi:\{(x_1,\cdots,x_{n-2},\psi(x_1,\cdots,x_{n-2},x_n),x_n),(x_1,\cdots,x_{n-2},x_n)\in\Omega\}\to\Omega\subset\mathbf{R}^{n-1}$，则 $\Psi\circ\Phi^{-1}:\omega\to\Omega$ 是迁移函数，其自变量是 $(x_1,\cdots,x_{n-1})$ 而函数值是 $(x_1,\cdots,x_{n-2},x_n)$．但是很清楚，因为

$$\Psi\circ\Phi^{-1}:(x_1,\cdots,x_{n-1})\to(x_1,\cdots,x_{n-2},x_n)$$

可以写为

$$x_j=x_j,\quad j=1,2,\cdots,n-2,$$
$$x_n=\varphi(x_1,\cdots,x_{n-1}),$$

且 $\dfrac{\partial x_n}{\partial x_{n-1}}=-\dfrac{\partial f}{\partial x_{n-1}}\Big/\dfrac{\partial f}{\partial x_n}\neq 0$，所以它当 $\omega,\Omega$ 充分小时是微分同胚，总之上述超曲面是一个 $n-1$ 维微分流形．

将它推广，在 $\mathbf{R}^n$ 中由 $m\ (m<n)$ 个方程

$$f_j(x_1,\cdots,x_n)=0,j=1,\cdots,m \tag{5}$$

定义的几何轨迹是一个 $n-m$ 维微分流形，这里 $f_j$ 是 $C^\infty$ 函数，而且 $m\times n$ 矩阵

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

之秩为 $m$.

这个例子告诉我们，$\mathbf{R}^n$ 中的曲面(5)是微分流形. 反过来，也有定理说明任一个 $k$ 维微分流形都可以嵌入在充分高维的欧氏空间 $\mathbf{R}^{2k+1}$ 中. 这就是说，为了研究微分流形，只需讨论本例就够了. 这时，$\mathbf{R}^n$ 成了此流形的包含空间. 但是我们并不采取这种观点，因为它使得许多很深刻的思想得不到充分表现. 而且，维数很高的空间带来的几何复杂性不见得比不用包含空间来抽象地讨论微分流形更少.

**例3　黎曼球面和射影空间**　上一章中我们在 $z$ 平面 $\mathbf{C}$(即 $(x,y)$ 平面 $\mathbf{R}^2$)上引进了一个无穷远点，并通过球极射影把它变成一个球面 $S^2$，称为黎曼球面. 从球极射影的公式可以看出这个射影其实是可微的. 现在我们要证明黎曼球面是一个 2 维微分流形，而球极射影其实就是坐标映射.

由图 7-2-1 可以看出，$S^2 = U_1 \cup U_2$，$U_1 = S^2 \setminus \{N\}$，$N$ 是北极 $(0,0,1)$；$U_2 = S^2 \setminus \{S\}$，$S$ 是南极 $(0,0,-1)$. 在 $U_1$ 中作球极射影 $\varphi_1: U_1 \to \mathbf{R}^2$，$\mathbf{R}^2$ 是过球心的赤道平面，并把一点 $P(\xi,\eta,\zeta)$，$\zeta \neq 1$ 映到 $\mathbf{R}^2$ 上的 $(x,y)$，很容易看到

$$x = \frac{\xi}{1-\zeta},\ y = \frac{\eta}{1-\zeta},\ \zeta \neq 1.$$

这就是 $U_1 \to \mathbf{R}^2$ 的坐标映射. 同样，若从南极 $S$ 开始作球极射影，则点 $(\xi,\eta,\zeta)$ 对应的点 $(x_1,y_1)$ 应为

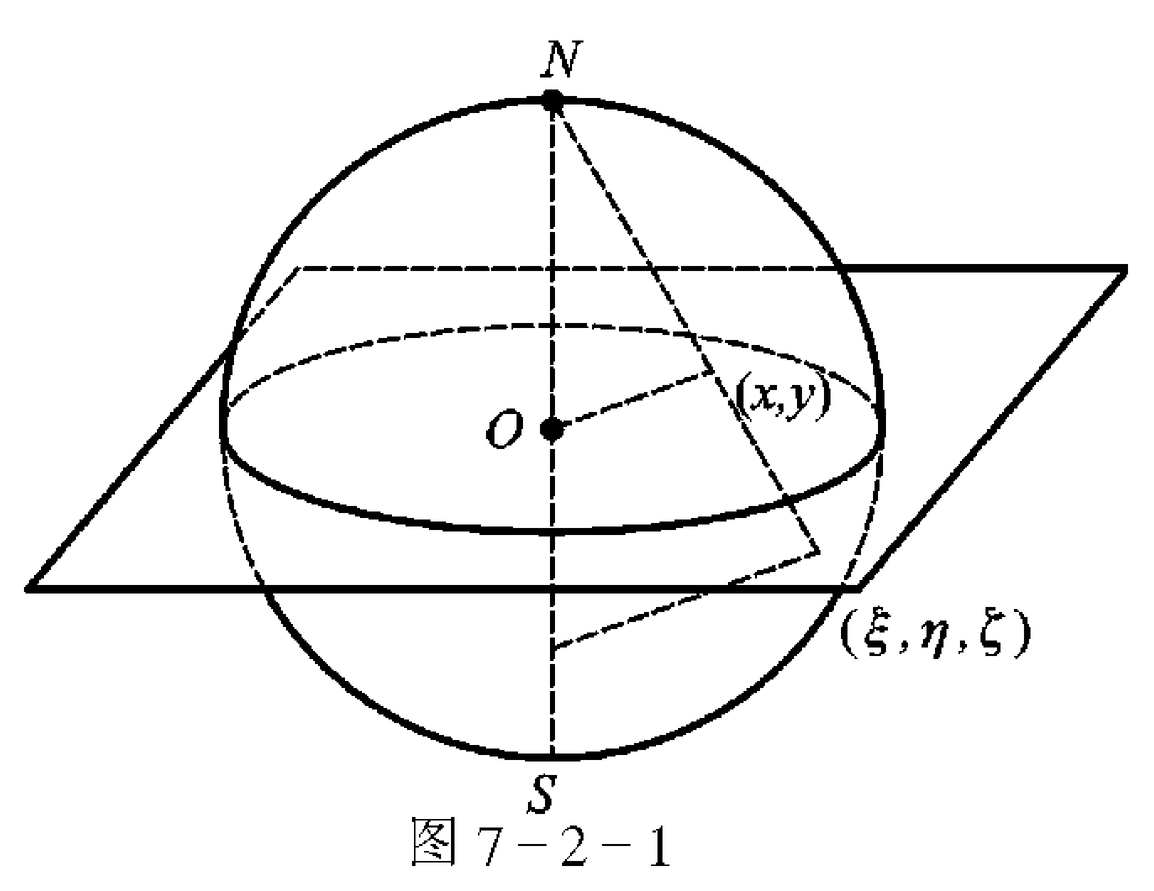


图 7-2-1

$$x_1 = \frac{\xi}{1+\zeta},\ y_1 = \frac{\eta}{1+\zeta},\ \zeta \neq -1.$$

这就是 $U_2 \to \mathbf{R}^2$ 的坐标映射. 由此就可以得到，除在 $(x,y) = (0,0)$ 处以外.

$$x_1 = \frac{x}{x^2+y^2},\ y_1 = \frac{y}{x^2+y^2}.$$

而在 $U_1 \cap U_2$ 中恰好没有南北极，因此不会有 $(x,y) = (0,0)$ 和 $(x_1,y_1) = (0,0)$. 所以上式就是 $U_1 \cap U_2$ 中的迁移函数. 由此我们看到，黎曼球面是一个微分流形. 其实，黎曼球面只不过是普通的复平面即 $(x,y)$ 平面 $\mathbf{R}^2$ 添加一个无穷远点而已，但是它却成了一个与 $S^2$ 同胚的球面，与 $\mathbf{R}^2$ 的性质大相径庭：前者是紧的，而后者只是局部紧，尽管局部地看来都是通常的平面. 如果我们进一步再看射影平面 $RP^2$，还会感受到更多的区别.

射影平面其实也只是普通平面再添加一些理想的元素. 这一次是添加了一条无穷远直线却又使它们的性质与黎曼球面大不相同了. 然而 $RP^2$ 仍然只是一个 2 维微分流形.

从上一章看到，定义 $RP^2$ 有多种方法，其一是在 $\mathbf{R}^3$ 中把过原点的直线当成一个元素，每一条这样的直线可以用其上非原点的一个点 $(x_1,x_2,x_3)$ 来定义，这里 $x_1,x_2,x_3$ 不能同时为 0. 不

过，若$(x_1, x_2, x_3)$与$(x'_1, x'_2, x'_3)$只相差一个常数因子，即

$$x_i = \lambda x'_i, \lambda \neq 0 \text{ 或 } x'_i = \mu x_i, \mu \neq 0$$

这两点要算成等价的：$(x_1, x_2, x_3) \sim (\lambda x_1, \lambda x_2, \lambda x_3)$. 而这些直线的集合即 $\mathbf{R}^3$ 中这些点的等价类的集合. 我们就称 $\mathbf{R}^3/\sim$ 为 2 维的实射影平面：$RP^2 = \mathbf{R}^3/\sim$. 上一章里我们在 $RP^2$ 上这样来定义一个拓扑，使 $RP^2$ 与球面 $S^2$ (经过一些改造)同胚. 现进一步问，能否用类似的方法在 $RP^2$ 上引入微分结构，使得在其上可以建立起微分学？

上一章里，我们是令通过原点的直线与 $S^2$ 相交，而将直线之集合映到 $S^2$ 上. 但每一条这样的直线与 $S^2$ 有两个交点互为对径点，因此要把 $S^2$ 上的对径点视为一点，这样得到射影平面的几何形象，也就是上文说的 $S^2$"经过一点改造"的意思. 这样作会发现，$RP^2$ 与一些很奇特的几何图形例如默比乌斯带有关. 也可以这样去讨论 $RP^2$ 的微分构造，但这需要把商拓扑的概念和方法讲得更明确，因此我们采用另一个方法.

讲解射影平面最简单的方法是引入 $\mathbf{R}^2$ 上的所谓"齐次坐标". 我们先引入 $\mathbf{R}^3$，即在 $\mathbf{R}^2$ 上方多加一个 $z$ 轴，于是 $\mathbf{R}^2$ 就成了 $z=0$. 我们先把这个平面向上抬起成为 $z=1$，于是原来的$(x, y)$点变成$(x, y, 1)$. 把它与原点$(0,0,0)$连成直线，这样一来，$\mathbf{R}^2$ 与 $\mathbf{R}^3$ 中过原点的直线族对应起来. 上面已说过这个直线可以用$(x, y, z)$($x, y, z$ 不同时为 0)之等价类$(x, y, z) \sim (\lambda x, \lambda y, \lambda z)$，$\lambda \neq 0$ 对应起来. 于是$(x, y)$点也就与$(x, y, 1)$之等价类对应起来了. 我们选此等价类之任一元$(x_1, x_2, x_3)$作为代表元，但要求 $x_1, x_2, x_3$ 不同时为 0，例如$(x, y)$对应于$(x, y, 1)$的等价类，而 $x_1 = x, x_2 = y, x_3 = 1$，于是 $x = x_1/x_3, y = x_2/x_3$. 正如复平面变成黎曼球面是通过附加无穷远点(即北极)而得，我们现在也对 $\mathbf{R}^2$ 添加"许多"点成一个新的集合，即加上$\{(x_1, x_2, 0)\}$. 并把这个新集合称为射影平面 $RP^2$. 它的点是坐标$(x_1, x_2, x_3)$(坐标不同时为 0)，称为射影平面中一点的齐次坐标. 当 $x_3 \neq 0$ 时，由它们可以得出普通的笛卡儿坐标 $x = x_1/x_3, y = x_2/x_3$. $RP^2$ 中每个点并不是恰有一个齐次坐标，而是有齐次坐标的恰好一个等价类. 齐次坐标方法很早在数学中就通行了. 例如当我们研究一个多项式 $P(x, y) = 0$ 时，就可以引入齐次坐标，并考虑一个三个变元的齐次多项式.

$$x_3^m P\left(\frac{x_1}{x_3}, \frac{x_2}{x_3}\right) = F(x_1, x_2, x_3) = 0.$$

$m$ 是 $P$ 的次数. 齐次多项式的好处不仅在于形状整齐，而且它允许添加一些理想的元素. 例如直线 $ax + by + c = 0$，现在变成 $ax_1 + bx_2 + cx_3 = 0$. 这样一来就增加了一个研究对象 $x_3 = 0$，它相应于 $a = b = 0, c \neq 0$. 本来我们是不应该考虑 $x_3 = 0$ 的，因为 $a = b = 0$ 而 $c \neq 0$ 与 $ax + by + c = 0$ 是矛盾的. 上面用 $x_3^m$ 去乘 $P\left(\frac{x_1}{x_3}, \frac{x_2}{x_3}\right)$ 也是不合法的. 但是若我们允许这样做了就相当于把集合$\{(x_1, x_2, 0)\}$即 $x_3 = 0$ 也纳入我们的讨论之中，就相当于把 $\mathbf{R}^2$ 变成 $RP^2$. 这种做法的合理性在于：$x_3 = 0$ 和 $ax_1 + bx_2 + cx_3 = 0$ 一样是一个一次方程，而一次方程代表一条直线，所以我们其实是在 $\mathbf{R}^2$ 中添上了一条直线——无穷远直线. 这条直线经过一个适当的射影变换即齐次坐标的非奇异线性变换，可以化为任一指定的普通的直线. 所以从射影几何的角度来看 $x_3 = 0$ 与其它的直线并无本质的不同.

现在我们可以来证明 $RP^2$ 确实是一个微分流形. 事实上

$$RP^2 = \bigcup_{i=1}^{3} U_i = \bigcup_{i=1}^{3} \{(x_1, x_2, x_3), x_i \neq 0\} \tag{6}$$

$U_i = \{(x_1, x_2, x_3), x_i \neq 0\}$显然是一个开集. 从 $U_i$ 到 $\mathbf{R}^2$ 之同胚定义为 $\varphi_i: U_i \to \mathbf{R}^2, (x_1, x_2, x_3) \mapsto \left(\frac{x_1}{x_i}, \cdots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \cdots, \frac{x_3}{x_i}\right)$，当 $i=1$ 或 3 时，$x_{i-1}$或 $x_{i+1}$无意义，这时要稍作些修改，总之使分子上不再出现 $x_i$. 余下的只要证明 $\varphi_i \circ \varphi_j^{-1}$ 在 $\varphi_j(U_i \cap U_j)$上是微分同胚即可. 不妨看 $\varphi_1 \circ \varphi_2^{-1}$. 在 $U_1 \cap U_2 (\neq \varnothing)$上，$\varphi_1(U_1 \cap U_2) = \left\{\left(\frac{x_2}{x_1}, \frac{x_3}{x_1}\right): x_1 x_2 \neq 0\right\} = \{(z_1, z_2)\}$，$\varphi_2(U_1 \cap U_2) = \left\{\left(\frac{x_1}{x_2}, \frac{x_3}{x_2}\right); x_1 x_2 \neq 0\right\} = \{(w_1, w_2)\}$，于是 $\varphi_1 \circ \varphi_2^{-1}$ 可以写为

$$w_1 = \frac{1}{z_1}, w_2 = \frac{z_2}{z_1}.$$

但因在 $U_1 \cap U_2$ 中 $z_1 = \frac{x_2}{x_1} \neq 0$，所以上式很明显是微分同胚. 于是 $RP^2$ 是一个微分流形.

比较一下黎曼球面与 $RP^2$，就知道，局部地看，它们是非常相似的，都是 $\mathbf{R}^2$. 其实凡微分流形必为局部欧氏空间. 但是整体来看却极不相同：$RP^2$ 虽然与黎曼球面同样为紧的、连通空间，却与后者不同，它是不可定向的. 一般地说来，$RP^n$ 当 $n$ 为偶数时必为不可定向的，而当 $n$ 为奇数时却是可以定向的. 那么 $RP^2$ 整体看来像什么样子？这就很不好说，因为它只能嵌入(这两个字也要解释，见后文)在 $\mathbf{R}^5$ 中. $\mathbf{R}^5$ 是一个什么样的空间就不是人们的直觉能够想到的了. 因此，整体地研究微分流形就得一整套其它方法. 这就是拓扑学的主题了. 它将应用同调、同伦等等一整套方法，而我们这一章只能介绍一点微分流形上的微积分学罢了.

**例 4 黎曼曲面** 本书中我们讲了许多不同的微分流形，不止于弯曲的曲面. 下面我们再讲两种在本书中涉及的微分流形. 其一是黎曼曲面. 在第二章中我们就复变量的多值函数 $w=\sqrt{z}$ 介绍了黎曼曲面的概念. 它是以我们的直觉为基础的，因而很难解释它如何穿过其自身. 现在我们从微分流形的角度来看 $w=\sqrt{z}$的黎曼曲面是如何构造出来的. 首先，我们把 $z$ 平面看成是 $\mathbf{R}^2$ 即$(x, y)$平面. 在其上取一点 $z_0 \neq 0$，并作以 $z_0$ 为心圆周过原点的圆盘 $U_0$，它是一个开集. 在这个开集中考查 $w=\sqrt{z}$. $\sqrt{z_0}$有两个可能的值，相差一个符号，任取其中之一，记作$(z_0)^{1/2}$，则有以下的二项级数展开式(见第三章)：

$$\begin{aligned} w = \sqrt{z} &= (z_0)^{\frac{1}{2}} \left(1 + \frac{z - z_0}{z_0}\right)^{\frac{1}{2}} \\ &= (z_0)^{\frac{1}{2}} \sum_{n=0}^{\infty} A_n \zeta_0^n, \qquad \zeta_0 = (z - z_0)/z_0. \end{aligned} \tag{7}$$

它在 $|\zeta_0| < 1$ 即当 $z \in U_0$ 时收敛. 因此(7) 是 $w = \sqrt{z}$ 在$U_0$ 中适用的表达式，或者说(7) 是代数方程 $w^2 - z = 0$ 的一个解. 另一解则是 $w = -(z_0)^{\frac{1}{2}} \sum_{n=0}^{\infty} A_n \zeta_0^n$. 我们在 $U_0$ 中任取一点 $z_1$，并用(7) 式算出

$$w = \sqrt{z_1} = (z_0)^{\frac{1}{2}} \sum_{n=0}^{\infty} A_n \left(\frac{z_1 - z_0}{z_0}\right)^n,$$

并记作$(z_1)^{\frac{1}{2}}$，于是又有

$$w = (z_1)^{\frac{1}{2}}\sqrt{1+\frac{z-z_1}{z_1}} = (z_1)^{\frac{1}{2}}\sum_{n=0}^{\infty}A_n\zeta_1^n,\ \zeta_1 = \frac{z-z_1}{z_1}, \tag{8}$$

这个表达式适合于圆盘 $U_1$(以 $z_1$ 为心，边界过原点)中(图 7-2-2)，而且由于$(z_1)^{\frac{1}{2}}$的选取方法知，在 $U_0\cap U_1$ 中，(7)与(8)定义相同的函数.第三章中指出，(8)称为(7)在 $U_1$ 中的解析延拓，于是由(7)与(8)共同定义的函数之定义域是 $U_0\cup U_1$，仿此进行下去，可能在作出 $U_{N-1}$时就发现 $z_0\in U_{N-1}$.于是我们可以选 $z_N$ 和 $U_N$ 即为 $z_0$，$U_0$.但这时会得到$(z_N)^{\frac{1}{2}}=-(z_0)^{\frac{1}{2}}$，这一点在第三章中就已讲过.所以，如果我们真的令 $U_N=U_0$，则在其上会定义两个函数，一是(7)式，另一是(7)的反号，因此无法在$\bigcup\limits_{i=0}^{N}U_i$上定义一个单值的解析函数 $w=\sqrt{z}$.为此，我们只好取另一个区域为 $U_N$.尽管它作为 $\mathbf{R}^2$ 的区域仍与 $U_0$ 相同，而且 $z_N=z_0$，却是另一个区域，并且从 $U_N$ 开始再用级数

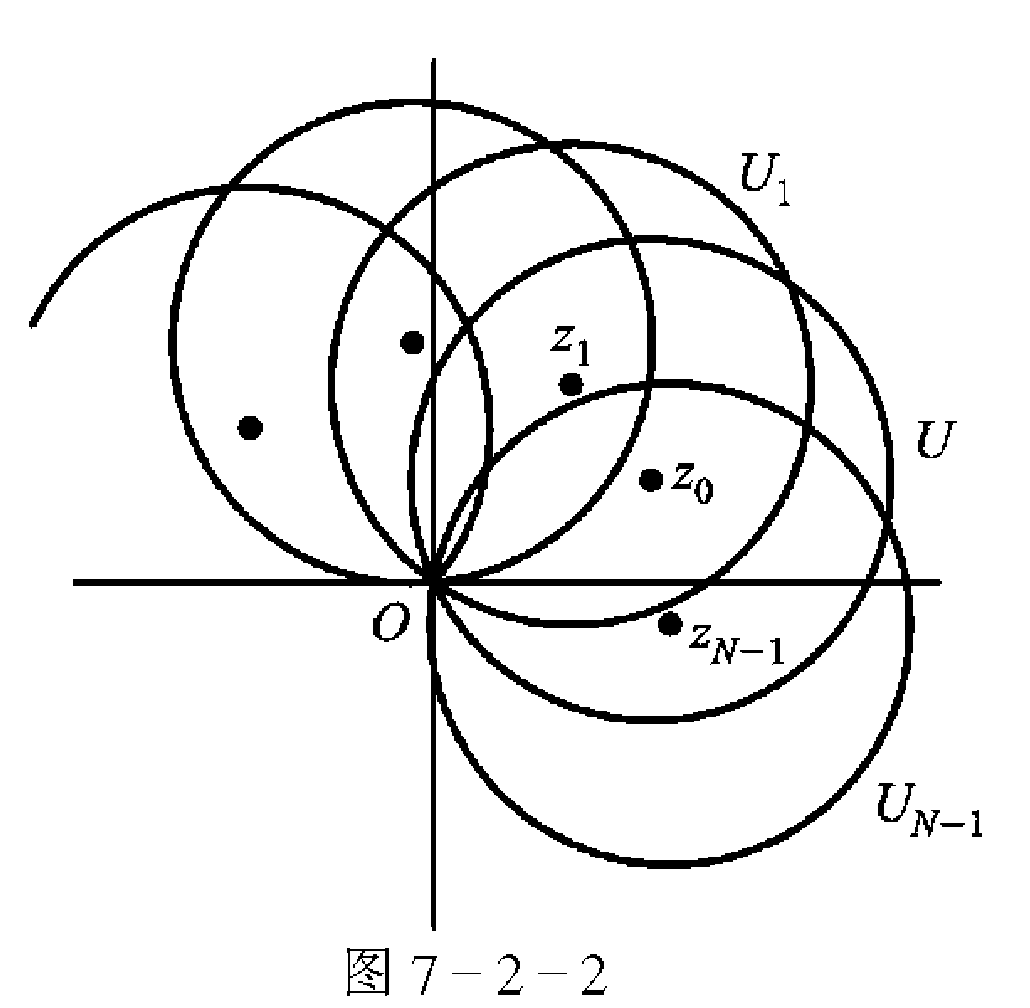


图 7-2-2

$$w = (z_N)^{\frac{1}{2}}\sum_{n=0}^{\infty}A_n\zeta_N^n,\ \zeta_N = \frac{z-z_N}{z_N} \tag{9}$$

作为 $w=\sqrt{z}$在 $U_N$ 中的定义(注意(7)与(9)的系数 $A_n$ 是相同的二项系数).如果这样作下去，则在作出了 $U_{2N-1}$时，就会发现 $z_{2N}=z_N=z_0$，但这时$(z_{2N})^{\frac{1}{2}}=-(z_N)^{\frac{1}{2}}=(z_0)^{\frac{1}{2}}$，于是我们又回到了出发点.这样做下去，可能得到许多 $U_j$ 而令 $M=\bigcup\limits_j U_j$，则 $w=\sqrt{z}$是定义在$M$ 上的单值函数.在 $M$ 中可能有某些$U_i$ 与 $U_j$ 适条件 $U_i\cap U_j\neq\varnothing$，例如 $U_{2N-1}\cap U_0\neq\varnothing$，$U_0\cap U_1=U_{2N}\cap U_{2N+1}\neq\varnothing$，但是 $U_0\cap U_N=\varnothing$，这样的做法自然与微分流形的定义十分相近了，只差迁移函数还没有做出来.

如果 $U_i\cap U_j\neq\varnothing$，则在 $U_i\cap U_j$ 中因为局部坐标$\zeta_i$ 与$\zeta_j$ 适合

$$\zeta_i = \frac{z-z_i}{z_i},\ \zeta_j = \frac{z-z_j}{z_j},\ z_i\neq 0,\ z_j\neq 0,$$

所以

$$\zeta_i = \frac{z_j}{z_i}(1+\zeta_j)-1,\ \zeta_j = \frac{z_i}{z_j}(1+\zeta_i)-1. \tag{10}$$

这就是我们需要的迁移函数.分开 $\zeta_i$ 与$\zeta_j$ 的实虚部，容易看到，上式是由 $\mathbf{R}^2$ 到 $\mathbf{R}^2$ 的微分同胚.但是更值得注意的是，作为复变量 $\zeta_i$ 与$\zeta_j$ 的关系来看，(10)表示 $\zeta_i$ 是$\zeta_j$ 的全纯函数，其逆亦然.仿照这样的方法我们可以研究代数方程

$$f(z,w)=0$$

(所谓代数方程即指 $f$ 是$z$ 与$w$ 的多项式)的解 $w=w(z)$.这些解一般地是 $z$ 的多值函数.仿照这个特例，而可以定义一般的黎曼曲面如下：

**定义 3** 设 $M$ 为一个 2 维微分流形，$\{(U_\alpha, \varphi_\alpha)\}$是其一个图册，如果当 $U_\alpha \cap U_\beta \neq \varnothing$时，$\varphi_\beta \circ \varphi_\alpha^{-1}: \varphi_\alpha(U_\alpha) \to \varphi_\beta(U_\beta)$是黎曼球面上的区域 $\varphi_\alpha(U_\alpha)$中的全纯函数，则称 $M$ 为一黎曼曲面.

至此，我们看到一个全纯函数可以定义在一个微分流形上，只要它有黎曼曲面构造即可，而不必要定义在复平面或黎曼球面上. 因此也就不必问这个黎曼曲面是怎样穿过它自身的. 问这个问题其实是问这个黎曼曲面可否在 $\mathbf{C}$ 上"实现". 这是不可能的，因为一般说来，一个 2 维微分流形要 $\mathbf{R}^5$ 才能把它装得下——即把它嵌入在 $\mathbf{R}^5$ 中，由此也就可见，微分流形概念怎样扩大了我们的视野.

**例 5 某些矩阵类** 上一节中我们介绍了一些在数学与物理学中起重要作用的变换群，特别重要的有 $O(n)$，$SO(n)$，$U(n)$，$SU(n)$等等. 当时我们指出，它们都是重要的李群. 李群是既有微分构造的微分流形，又是有群构造的群，而且这两种构造应该是相容的(相容性的确切意义这里不能讲了). 现在我们来看一下这些变换群的微分流形构造.

首先，我们把这些变换都看或 $n \times n$ 矩阵. 于是每一个矩阵都对应于 $\mathbf{R}^{n^2}$ 中的一个点. 而一个特殊的变换群，将成为 $\mathbf{R}^{n^2}$ 中的一个流形. 如我们在例 2 中看到的那样，这些流形将由 $\mathbf{R}^{n^2}$ 中的很多个方程来定义. 例如 $SO(3)$是适合以下条件的 $3\times 3$ 矩阵，$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$定义的. 这里

$$\det A = 1, \tag{11}$$

$$\sum_{k=1}^{3} a_{ik}a_{jk} = \delta_{ij}, \quad i, j = 1, 2, 3. \tag{12}$$

问题是，这些条件是否互相独立的？又是不是足够的？例如上面这组条件是说 $A$ 之各行成为一个 o.n. 系，那么是不是还应增加关于各列的条件

$$\sum_{k=1}^{3} a_{ki}a_{kj} = \delta_{ij}, \quad i, j = 1, 2, 3$$

才够？事实上是不必要的，有关于行的(12)就够了. 但是即令在(12)式中还有多余的. (12)中有 9 个条件，而例如 $i=1, j=2$ 以及 $i=2, j=1$ 其实是一个条件. 所以这 9 个条件中实际上独立的最多只有 6 个. 再有 $\det A = 1$. 事实上，由(12)已知 $A \in O(n)$，从而 $\det A = \pm 1$，上一节已指出 $\det A = 1$ 表示在变换 $A$ 之下手征不变. 如果适合手征不变的正交变换构成一个流形，则改变手征的正交变换将成为另一个流形. 这样，$O(n)$其实是两个流形，而 $\det(A) = 1$ 只表示由这两个流形中先取了一个. 所以条件(11)并不是与(12)互相独立的第 7 个条件.

从这些分析可以看到，我们现在遇到的困难将是弄清究竟有多少个独立的条件. 这些问题在研究这些群时都是最起码的问题，我们在这里不去讨论它. 我们只来看一个最简单的特例，即证明 $SL(n, \mathbf{R})$是一个 $n^2 - 1$ 维微分流形，$SL(n, \mathbf{R})$称为特殊线性群，即适合 $\det A = 1$ 的 $n \times n$ 实矩阵所成之群. $GL(n, \mathbf{R})$称为一般线性群，即适合 $\det A \neq 0$ 的 $n \times n$ 实矩阵所成之群. 这个证明其实是很简单的.

把一个矩阵 $A$ 看成 $\mathbf{R}^{n^2}$ 中的一个点，则 $SL(n, \mathbf{R})$就是 $\mathbf{R}^{n^2}$ 中由方程 $\det A = 1$ 所定义的一个子集. $\det A$ 是 $n^2$ 个自变量 $a_{ij}$所成的一个很特殊的多项式. 要看它是否定义一个流形就要看

在其上一点 $a_{ij}$ 处(这些 $a_{ij}$ 适合方程 $\det A=1$)是否没有奇点,即$\frac{\partial}{\partial a_{ij}}(\det A)$是否不全为 0.但这是很容易弄清的,令 $a_{ij}$ 在 $A$ 中的代数余子式为 $A_{ij}$,显然其中不含 $a_{ij}$,容易算出

$$\frac{\partial}{\partial a_{ij}}(\det A)=0 \Longleftrightarrow A_{ij}=0, \forall i,j.$$

但若在一点处一切 $A_{ij}=0$,则在该点由拉普拉斯展式

$$\det A=\sum_{j=1}^{n} a_{ij}A_{ij}, \quad i \text{ 为固定的}$$

知 $\det A=0$,而这与 $\det A=1$ 相矛盾.

所以,$SL(n,\mathbf{R})$是一个 $n^2-1$ 维微分流形.关于 $SO(n)$将在下面讨论.

现在讨论微分流形间的映射.设 $M$ 和 $N$ 是两个微分流形,其维数分别为 $m$ 和 $n$,而 $f$ 是由 $M$ 到 $N$ 中的映射,且 $f(P)=Q$.为了定义映射的可微性,我们取 $P$ 和 $Q$ 的坐标邻域 $U$ 与 $V$ 以及相应的坐标映射 $\varphi$ 和 $\psi$.于是 $\psi\circ f\circ\varphi^{-1}:\varphi(U)\to\psi(V)$是由 $\mathbf{R}^m$ 中的一个区域 $\varphi(U)$到 $\mathbf{R}^n$ 中的区域 $\psi(V)$的映射,在局部坐标下就是一组 $n$ 个 $m$ 元函数.

**定义 4**　设 $M,N,f$ 等均如上所述,若任一点 $P\in M$ 以及 $Q=f(P)\in N$ 均有坐标邻域 $U$ 与 $V$ 以及相应坐标映射 $\varphi,\psi$ 使 $\psi\circ f\circ\varphi^{-1}:\varphi(U)\subset\mathbf{R}^m\to\psi(V)\subset\mathbf{R}^n$ 为 $C^\infty$ 光滑函数,则称 $f$ 为 $C^\infty$ 可微映射.

这个定义与局部坐标的选择无关,因为设有另一组坐标邻域与坐标映射$(\varphi_1,U_1)$,$(\psi_1,V_1)$则在 $\varphi_1(U_1\cap U)$中,$\psi_1\circ f\circ\varphi_1^{-1}=(\psi_1\circ\psi^{-1})\circ(\psi\circ f\circ\varphi^{-1})\circ(\varphi_1\circ\varphi^{-1})^{-1}$是光滑的,因此 $f$ 仍为可微映射.

以上我们结束了有关微分流形及其上的映射的基本概念的讨论,下面就开始讨论微分流形上的微分学问题.

**2. 切空间、余切空间以及切映射**　微分学的基本思想是把研究对象线性化.线性化以后出现性质比较简单的对象,然后可以从它们的性质推断原来对象的性质.例如,研究一条曲线,一个曲面时,我们就用它们在某点附近的切线或切面去局部地取代它们,并进而讨论曲线与曲面的性质.同样,研究一个函数(也就是一个映射)时,就用其微分去取代它,以研究它的性质.所以,在讨论微分流形上的微分学(积分学的讨论见最后一节)时,我们也就是要看怎样发展上面讲的基本思想.

于是,设有 $n$ 维微分流形$M$ 以及一点$P\in M$.首先看如何推广 $M$ 在$P$ 点的切面.当然,这个切面其实是由 $M$ 上过$P$ 点的曲线之切线组成的.但是,$M$ 中一条曲线过某点的切线一般会离开这一点而伸向 $M$ 以外的地方去的,于是要问,切线的概念是否本质地依赖于包含空间?这个问题很容易从力学上回答.因为一个质点沿其轨道运动时,其速度是一个向量,其大小表示速度之大小,而其方向则是切线的方向.所以,速度向量也可以看作一个切向量.如果两个质点都以同样的速度经过 $P$ 点,则除了很特殊的发生奇异性的情况以外,它们的轨道将在 $P$ 点相切,而这些轨道在 $P$ 点就有了相同的切向量,也就有了相同的切线."相切于某点 $P$"这个关系是一个等价关系:任一曲线 $c$ 必在$c$ 上某一点$P$ 处与其自身相切(自反性);若 $c_1$ 在 $P$ 点切于$c_2$,则反过来 $c_2$ 也在 $P$ 处切于$c_1$(对称性);若 $c_1$ 在 $P$ 点切于$c_2$,$c_2$ 在 $P$ 点又切于$c_3$,则 $c_1$ 在 $P$ 点切于$c_3$(传递性).于是我们可用这种等价关系将过点的曲线 $c$ 分成等价类,并且用这种等价关系代替切向量.

甚至可以说,所谓切向量就是这样的等价类.看切线是怎样组成切平面的,就是看这些切向量如何构成一个线性空间,其维数是否与曲面相同.想要这些等价类构成线性空间,就需要定义这些等价类的线性运算.第三章提到了,这些等价类确实构成一个 $n$ 维空间,但未详细解释,现在讨论如下:现在我们没有了包含空间,但是微分流形在其每一点附近,在某一个坐标邻域中都与 $\mathbf{R}^n$ 同胚,我们就要利用这种同胚性来回答以上问题.下面着手具体讨论它,这时请注意可微性的作用.

设 $c: I\to M$ 为一光滑曲线.这里 $I=(-\varepsilon,\varepsilon)$,$\varepsilon>0$.$c(0)=P\in M$,令 $U$ 为 $P$ 的一个坐标邻域,$\varphi$ 为相应的坐标映射,于是 $\varphi\circ c: I\to\mathbf{R}^n$ 可以写为 $(x^1(t),\cdots,x^n(t))$,是 $t$ 的光滑向量,而有切向量 $\xi(t)=(\dot{\varphi}\circ c)(t)=(\xi^1(t),\cdots,\xi^n(t))$ 存在,这里 $\xi^i(t)=\dot{x}^i(t)$.如果有两条这样的过 $P$ 的曲线 $c_1(t)$,$c_2(t)$ 使得对应的 $\xi_{(i)}(t)$,$i=1,2$,适合 $\xi_{(1)}(0)=\xi_{(2)}(0)$,就说 $c_1(t)$ 与 $c_2(t)$ 在 $t=0$ 处,即在 $P$ 点相切.相切关系是一个等价关系,记作 $\simeq$,$c$ 的等价类记作 $[c]$,于是称 $[c]$ 为 $M$ 在 $P$ 点处的一个切向量,$\xi(0)$ 就是这个切向量之坐标表示.如果 $M$ 在 $P$ 点有另一个局部坐标 $(V,\psi)$,于是 $\varphi\circ c$ 现在变成 $\psi\circ c=(y^1(t),\cdots,y^n(t))$,而 $\xi(t)$ 变成 $\eta(t)=(\dot{y}^1(t),\cdots,\dot{y}^n(t))$.$\eta(0)=(\dot{y}^1(0),\cdots,\dot{y}^n(0))$ 是 $M$ 在 $P$ 点处的切向量之另一坐标表示.容易看到相切关系在新坐标系下仍得以保持,事实上

$$\frac{\mathrm{d}}{\mathrm{d}t}(\psi\circ c)\Big|_{t=0}=\frac{\mathrm{d}}{\mathrm{d}t}[(\psi\circ\varphi^{-1})\circ(\varphi\circ c)]_{t=0}=\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}\Big|_P(\dot{\varphi}\circ c)\Big|_{t=0}. \tag{13}$$

所以等价类 $[c]$ 并不因采用不同的局部坐有而有所不同.而且因为 $\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}\Big|_P=\frac{\partial y}{\partial x}\Big|_P$ 是迁移函数之雅可比矩阵,所以是非奇异的.切向量如果能构成一个线性空间,则不同的坐标表示成为互相同构的线性空间,称为 $M$ 在 $P$ 点处的切空间,记为 $T_PM$.所以我们现在要做的是弄清 $T_PM$ 的线性结构.

为此,设有 $[c_1]$,$[c_2]\in T_PM$,$\lambda_1,\lambda_2\in\mathbf{R}$.我们要问,如何定义 $\lambda_1[c_1]+\lambda_2[c_2]$,为此取 $[c_1]$,$[c_2]$ 的代表元 $c_1$ 与 $c_2$,而得到 $\mathbf{R}^n$ 中的两条曲线 $\varphi\circ c_1$ 与 $\varphi\circ c_2$ 以及切向量之坐标表示 $\xi_{(1)}\Big|_{t=0}=(\dot{\varphi}\circ c_1)\Big|_{t=0}$ 与 $\xi_{(2)}(t)\Big|_{t=0}=(\dot{\varphi}\circ c_2)\Big|_{t=0}$.因为它们都是 $\mathbf{R}^n$ 中的向量,所以 $c_1\xi_{(1)}(0)+c_2\xi_{(2)}(0)$ 是有意义的.而且因为有(13)式,所以这里的代数运算的结果不因采用不同的局部坐标而有异.现在我们要证明

**定理 2** $T_PM$ 是一个 $n$ 维空间.

**证** 上面已看到 $T_PM$ 中的之坐标表示之间确可定义线性运算.如果 $\zeta\in\mathbf{R}^n$,则一定可以找到一条曲线 $c$ 使 $\zeta$ 恰好是它 $[c]$ 的坐标表示.事实上,若设 $\varphi: P\mapsto 0$($\mathbf{R}^n$ 中的原点)则 $x(t)=\zeta t$ 是 $\mathbf{R}^n$ 中过 $P$ 的一条以 $\zeta$ 为切向量的直线.$\varphi^{-1}(x(t))$ 则是 $M$ 中过 $P$ 的曲线,而且 $\zeta$ 恰好是它的切向量之坐标表示.即是说任给 $\mathbf{R}^n$ 中一个向量,必可找到 $M$ 中一个过 $P$ 之曲线的等价类,以此向量为切向量.于是我们可以在曲线的等价类中定义线性运算 $\lambda_1[c_1]+\lambda_2[c_2]$ 即以 $\lambda_1\xi_{(1)}+\lambda_2\xi_{(2)}$ 为坐标表示的等价类.这个等价类是一定存在的,因为我们已找出了它的一个代表元:$\varphi^{-1}((\lambda_1\xi_{(1)}+\lambda_2\xi_{(2)})t)$.于是 $T_PM$ 成了一个线性空间,余下的只要证明它的维数为 $n$ 即可.

把上面的证明综合成一句话，就是我们可以通过坐标映射把 $\mathbf{R}^n$ 中的线性关系移到 $T_PM$ 中来，而且因为有(13)式可知这种线性关系的转移不受局部坐标选择的影响. 特别是 $0\in\mathbf{R}^n$ 所对应的曲线之等价类我们就应该规定为 $T_PM$ 中的零元素. 由于这个等价类之重要性，我们要特别看一下，其中包括了什么. 当然，其中包含了 $\varphi^{-1}(0\cdot t)$. 这里的 0 是零向量而不是数 0，即 $0=(0,\cdots,0)$，所以我们写 $0\cdot t$ 而不写 $0t$. $\varphi^{-1}(0\cdot t)$ 是这样的曲线：即对一切 $t$ 值它始终是一个点 $P$，我们不妨称之为驻定曲线. 但除此以外，它还包含了 $\mathbf{R}^n$ 中的适合 $\dot{x}(0)=0$ 的曲线在 $\varphi$ 下的原像，而且也只包含这种曲线. 所以 $[c]=\{\varphi^{-1}(x(t)),\dot{x}(0)=0\}$. 但是若一曲线 $x=x(t)$ 适合 $\dot{x}(0)=0$ 就说它以 $t=0$ 为奇点. 因此 $T_PM$ 中的零元素即是以 $P$ 为奇点的曲线之等价类. 这一等价类我们记为 $[0]$. 有了这些规定后即得

$$\sum_{i=0}^{k}\lambda_i\xi_{(i)}=0\Longleftrightarrow\sum_{i=1}^{k}\lambda_i[c_i]=[0].\tag{14}$$

由此，$\{[c_i]\}$，$i=1,2,\cdots,k$，在 $T_PM$ 中线性相关或线性无关的充分必要条件是 $\{\xi_{(i)}\}$，$i=1,2,\cdots,k$，在 $\mathbf{R}^n$ 中线性相关或线性无关. 但是 $\mathbf{R}^n$ 中恰好有而且最多也只有 $n$ 个线性无关的向量，所以在 $T_PM$ 中也是这样. 由此可知 $T_PM$ 是 $n$ 维空间，证毕.

于是要问，如何找出 $T_PM$ 的一个基底？为了找出一个最有用的基底，我们要再回顾一下切空间的定义. 切向量和切空间在 $\mathbf{R}^n$ 中都是很直观的东西，但是当微分流形 $M$ 并未放置在某包含空间中时，就无法应用这些直观的东西. 于是我们利用微分流形局部地与欧氏空间同胚这一事实，并只在一个坐标邻域 $U$ 中考虑切向量与切空间，而且用坐标映射 $\varphi$ 把我们的考察引入 $\varphi(U)\subset\mathbf{R}^n$. 在 $\mathbf{R}^n$ 中则可以充分应用直观上已经很明确的概念，然后用 $\mathbf{R}^n$ 的线性同构，作出抽象的 $T_PM$. 这样做法还给了我们更大的活动余地：简单地说，只要与 $\varphi(U)\subset\mathbf{R}^n$ 上的直观的切空间同构的就算是 $M$ 在 $P$ 点的切空间 $T_PM$. 上面我们用速度向量来构成线性空间. 现在我们使用 $\mathbf{R}^n$ 上的方向导数来建立切空间理论. 先讨论 $M=\mathbf{R}^n$ 的情况.

首先，$\mathbf{R}^n$ 中的方向导数是一个求导运算，这是因为，若 $F(x)$ 是定义在某点 $P\in\mathbf{R}^n$ (不妨设 $P$ 就是 $x=0$) 附近的光滑函数，则在某一个方向 $\xi$ ($\xi\in\mathbf{R}^n$ 是一个向量) 上在 $P$ 点对 $F(x)$ 求导，就是把此函数放在过 $P$ 的直线 $x=\xi t$ ($t=0$ 对应于 $P$ 点) 上，得到一个 $t$ 的函数 $F(x)=F(\xi t)$，然后对 $t$ 求导，得到

$$X(F)=\frac{\mathrm{d}}{\mathrm{d}t}F(\xi t)=\sum_{i=1}^{n}\xi^i\frac{\partial F(\xi t)}{\partial x^i}=\langle\mathbf{grad}\ F(\xi t),\xi\rangle.\tag{15}$$

(注意，我们这里写的是 $\xi^i$ 而不是 $\xi_{(i)}$，上面(14)式讲 $\xi$ 与 $[c]$ 的对应关系时则用的是记号 $\xi_{(i)}$. 那里是表示的第 $i$ 个向量，现在则表示向量 $\xi$ 的第 $i$ 个分量，分量记号用上标是否指 $\xi$ 为逆变向量？是的，这一点下面再说.) 最后令 $t=0$ 得到在 $P$ 点的方向导数值. 这里有一个非常关键的事实，如果我们不把 $F$ 放在直线 $x=\xi t$ 上，而且放在该直线所属的等价类 $[c]$ 中任一曲线上，所得结果全是一样的. 所以我们所建立的不只是 $X$ 与 $x=\xi t$ 的对应关系，而且是 $X$ 与 $[c]$ 的对应关系. (15)式中的算子 $X$ 有以下的性质：

(1) 关于 $F$ 的线性性质，即对光滑函数 $F,G$ 和实数 $\lambda,\mu$ 有

$$X(\lambda F+\mu G)=\lambda X(F)+\mu X(G).\tag{16}$$

(2)若 $F=\text{const}$,则

$$X(F)=0.$$

(3)莱布尼茨性质,即

$$X(FG)=FX(G)+GX(F). \tag{17}$$

特别值得注意的是(17).事实上,性质(2)可由(16),(17)导出.凡一线性算子具有以上三个性质的都称为一个导子(derivation),我们以后还会看到并不是只有方向导数才是导子,而例如微分也是导子.方向导数在导子中的特点在于,它以某种方式与一个向量相联系.为了看清这种联系,注意到上面讨论中的 $P$ 本来可以在 $\mathbf{R}^n$ 中某一区域上变动,现在我们把 $P$ 固定只看 $\xi\in\mathbf{R}^n$ 变动,则(15)式表示当 $P$ 固定(即是令 $t=0$ 时)有一个由 $\xi\in\mathbf{R}^n$ 到某一个 $X$ 的映射.这些 $X$ 都可以写为

$$X_P(F)=\sum_{i=1}^n\xi^i\frac{\partial F(0)}{\partial x^i}. \tag{18}$$

它与(15)之区别在于 $F$ 之导数是在 $P$ 即 $t=0$ 处取值的,所以我们用 $X_P$ 表示,以示与(15)的区别.这个区别似乎是微妙的,但是它是下面将要讨论的切丛概念的基础.

**定理 3** 由(18)式定义的映射赋予 $\{X_P\}$ 以线性构造,而且使 $\mathbf{R}^n$ 与 $\{X_P\}$ 同构.

**证** $\{X_P\}$ 的线性构造是明显的,因为

$$\lambda\xi+\mu\eta\Rightarrow\sum_{i=1}^n(\lambda\xi^i+\mu\eta^i)\frac{\partial F(0)}{\partial x^i}.$$

这里 $\xi,\eta\in\mathbf{R}^n$ 是向量,$\xi=(\xi^1,\cdots,\xi^n)$,$\eta=(\eta^1,\cdots,\eta^n)$,而 $\lambda,\mu$ 是实数.从这一个简单的推导就看出为什么要把 $P$ 固定,因为若在计算中令 $P=\xi t$ 而让 $P$ 变动,计算的结果只会得出

$$\sum_{i=1}^n\lambda\xi^i\frac{\partial F(\xi t)}{\partial x^i}+\sum_{i=1}^n\mu\eta^i\frac{\partial F(\eta t)}{\partial x^i}, \tag{19}$$

而无法结合成 $F(\xi t+\eta t)$(即令合成了也会丧失了(18)式对 $\xi,\eta$ 的线性依赖性,(15)式对 $\xi,\eta$ 本来就不是线性的).

在给出了 $\{X_P\}$ 的线性构造以后,(18)式为一同构就容易证明了.首先它是单射,因为若 $\xi\in\mathbf{R}^n$ 对应的 $X_P$ 是零算子,则应当有:对任一光滑函数 $F(x)$,

$$\sum_{i=1}^n\xi^i\frac{\partial F(0)}{\partial x^i}=0,$$

依次令 $F=x^1,x^2,\cdots,x^n$ 就得出 $\xi^1=0,\xi^2=0,\cdots,\xi^n=0$,即 $\xi=0$.

最后证明(18)是满射.为此,先要注意一件事.若 $F(x)$ 在 $x=0$ 附近是光滑的,把它在 $x=0$ 处用泰勒公式展开到二次项:

$$F(x)=F(0)+\sum F_i(0)x^i+\sum_{i,j=1}^n F_{ij}(x)x^ix^j,$$

用 $X_P$ 作用到两边,这里 $F_i(0)=\dfrac{\partial F(0)}{\partial x^i}$,而 $F_{ij}(x)$ 仍是 $x=0$ 附近的光滑函数.利用 $X$ 之性质(1)~(3),再在 $P$ 点取值即有

$$X_P(F)=\sum_{i=1}^n F_i(0)X_P(x^i).$$

现在在 $\mathbf{R}^n$ 中取向量 $\xi=(\xi^1,\cdots,\xi^n)$,我们只要这样定义算子 $X$,使 $\xi=(X_P(x^1),\cdots,X_P(x^n))$ 即

$X=\sum_{i=1}^{n}\xi^i\frac{\partial}{\partial x^i}$，代入上式即知对任意 $F(x)$ 有(18)式

$$X_P(F)=\sum_{i=1}^{n}\xi^i\frac{\partial F(0)}{\partial x^i}.$$

所以任一个 $X_P$ 都是某一个 $\xi\in\mathbf{R}^n$ 之像，因此满射得证.

这个定理告诉我们 $T_PM$ (现在 $M$ 就是 $\mathbf{R}^n$) 与 $\{X_P\}$ 同构. 因此我们可以说 $\{X_P\}$ 就是 $T_PM$，或者说切向量就是导子 $X_P$，而且由(18)可见 $\left\{\frac{\partial}{\partial x^1},\cdots,\frac{\partial}{\partial x^n}\right\}$ 是 $T_PM$ 的一个基底，称为由局部坐标 $(x^1,\cdots,x^n)$ 诱导出的基底.

现在我们来看坐标变换下 $X_P$ 的变化. 注意到 $\xi$ 是速度向量 $\xi=(\xi^1,\cdots,\xi^n)=(\dot{x}^1(0),\cdots,\dot{x}^n(0))$，所以 $\xi$ 与 $x$ 一样是逆变向量. $\left(\frac{\partial F(0)}{\partial x^1},\cdots,\frac{\partial F(0)}{\partial x^n}\right)$ 就是古典的梯度算子，因此是协变向量. 所以当坐标由 $x$ 变为 $y$ 时，若 $y$ 坐标系下的速度为 $\eta=(\eta^1,\cdots,\eta^n)$，假设 $x=0$ 变为 $y=0$，则

$$\xi^i=\frac{\partial x^i}{\partial y^j}(0)\eta^j,$$

$$\frac{\partial F(0)}{\partial x^i}=\frac{\partial y^k}{\partial x^i}(0)\frac{\partial F(0)}{\partial y^k}.$$

把它们代入(18)，即得

$$X_P(F)=\sum_{i,j,k=1}^{n}\frac{\partial x^i}{\partial y^j}(0)\frac{\partial y^k(0)}{\partial x^i}\eta^j\frac{\partial F(0)}{\partial y^k}=\sum_{j,k=1}^{n}\delta_j^k\eta^j\frac{\partial F(0)}{\partial y^k}$$
$$=\sum_{j=1}^{n}\eta^j\frac{\partial F(0)}{\partial y^j}.$$

所以(18)定义的 $X_P$ 其实是与坐标无关的，所以说它是切向量就更有理了.

以上我们讲的是 $\mathbf{R}^n$ 的切向量与切空间，现在问应该如何对微分流形 $M$ 上一点 $P$ 定义 $T_PM$. 为此取 $P$ 的一个坐标邻域 $U$ 与坐标映射 $\varphi$，并用 $x$ 表示 $\varphi(U)\subset\mathbf{R}^n$ 中的坐标. 注意，直接用(18)式是有困难的，因为 $\frac{\partial}{\partial x^i}$ 中的 $x^i$ 是 $\varphi(U)$ 中的量而不是 $M$ 上的量. 所以在 $M$ 上 $\frac{\partial}{\partial x^i}$ 是不能定义的. 为此，我们仍按上面讲的基本思想来处理这个问题：把坐标邻域 $\varphi(U)$ 中的结果直接移到 $M$ 上 $P$ 点的邻域中去. 这时只要注意一个问题：$P\in M$ 可以有几个不同的坐标邻域. 例如 $(U,\varphi)$ 与 $(V,\psi)$，我们只需证明使用 $(U,\varphi)$ 与使用 $(V,\psi)$ 结果是一致的即可. 因此我们现在设 $F$ 是 $M$ 上的函数. 于是 $F\circ\varphi^{-1}$ 是 $\varphi(U)\subset\mathbf{R}^n$ 上的光滑函数，亦即为 $x$ 的光滑函数，使得 $\frac{\partial}{\partial x^i}(F\circ\varphi^{-1})$ 有意义. $\xi$ 本来是速度 $\dot{x}(0)$，也是与局部坐标相联系的，而且如前所述，$\left(\frac{\partial(F\circ\varphi^{-1})}{\partial x^1},\cdots,\frac{\partial(F\circ\varphi^{-1})}{\partial x^n}\right)$ 是 $\mathbf{R}^n$ 上的协变向量，$\xi\in\mathbf{R}^n$ 则是逆变向量，我们给出

**定义 5** 对上述 $M$，$F$ 与 $P\in M$，我们定义 $P$ 处的导子 $X_P$ 为

$$X_P(F)=\sum_{i=1}^{n}\xi^i\frac{\partial}{\partial x^i}(F\circ\varphi^{-1})(0). \tag{20}$$

这里右方的 0 即是 $\varphi(P)$，$X_P$ 称为 $M$ 在 $P$ 点的切向量.

这样定义的 $X_P$ 自然适合导子的三个要求.

这个定义显然与坐标的选取无关.若改用另一个坐标邻域$(V,\psi)$,而以 $y$ 为新的局部坐标,并设在新坐标 $y$ 下,逆变向量 $\xi$ 的分量变成 $\eta=(\eta^1,\cdots,\eta^n)$,且

$$\xi^i=\frac{\partial x^i}{\partial y^j}(0)\eta^j,$$

代入(18)式有

$$\begin{aligned}X_P(F)&=\sum_{i=1}^n\eta^j\frac{\partial x^i}{\partial y^j}(0)\frac{\partial(F\circ\varphi^{-1})}{\partial x^i}(0)\\&=\sum_{j=1}^n\eta^j\frac{\partial[F\circ\varphi^{-1}\circ(\varphi\circ\psi^{-1})]}{\partial y^j}(0)\\&=\sum_{j=1}^n\eta^j\frac{\partial(F\circ\psi^{-1})}{\partial y^j}(0).\end{aligned}$$

它的形状仍与(20)相同.由此可见这样定义的对象 $X_P(F)$确实仅仅决定于 $M$ 的性质(当然只是局部性质),而与坐标之选取无关.也由此,我们可以把定理 3 改写为

**定理 3′** 设 $M$ 为一 $n$ 维微分流形,$P\in M$.则由定义 5 给出的 $X_P$ 之集合具有自然的 $n$ 维线性空间结构,且$\{X_P\}\simeq\mathbf{R}^n$.$\{X_P\}$称为 $M$ 在 $P$ 点的切空间.记作 $T_PM$.

至此,一个微分流形 $M$ 在 $P$ 点附近的"线性化"过程完成了.我们在 $M$ 上 $P$ 点处给 $M$ 添加了一个 $n$ 维线性空间 $T_PM$.我们可以在某种意义下用 $T_PM$ 来代替 $M$ 以研究其性质.这样自然产生许多深刻的问题,例如:可否局部地由 $T_PM$ 恢复 $M$(指数映射)?如果 $M$ 还是一个李群,则群结构在 $T_PM$ 上诱导出什么(李代数)?所有这一切自然都远远超出本书范围.但是有一个问题却是不能回避的.

现在有了 $T_PM$,它是一个线性空间,线性空间上有许多结构,§1 讲的向量和张量就是其中十分重要的.有许多结构需要考虑 $P$ 在 $M$ 上的变化(而不是如上面那样把 $P$ 固定,而我们上面一再指出,把 $P$ 固定是很重要的一步),这就导致了向量丛的概念,我们等一会再去讨论它.目前要讨论的是 $T_PM$ 的对偶空间,但在此以前需要讨论什么是切映射.

首先我们要重新看一下什么是一个函数的微分,看一下一元函数这个最简单的例子(图 7-2-3).如果一个光滑函数 $f(x)$定义在区间$(a,b)$上,而过 $P$ 点有一条切线.设 $P$ 点的坐标是 $x$(在上文中是取 $x$ 为 0).我们不妨把函数的图形看成一个微分流形 $M$,而视此函数关系为由一维微分流形 $\mathbf{R}$ 到 $M$ 的映射 $f$,映 $Q\in\mathbf{R}$ 到 $P\in M$,而且此映射是可微的.$M$ 在 $P$ 点的切空间就是过 $P$ 的切线,而 $\mathbf{R}$ 在 $Q$ 点的切空间仍是一个 1 维线性空间 $\mathbf{R}$,不过原点要放在 $Q$ 点:$T_Q\mathbf{R}=\mathbf{R}$.当我们有一个可微映射 $f:(a,b)\subset\mathbf{R}\to M,x\mapsto f(x)$时,在切空间 $T_Q\mathbf{R}$ 与 $T_PM$ 之间也有一个映射 $A$,映 $T_Q\mathbf{R}$ 中的 $h$ 为 $Ah=f'(x)h$.不过我们并没有把它画在 $T_PM$ 上,而是画在与 $y$ 轴平行的方向上,因为这是我们习惯的.通常的微积分教本把 $Ah$ 这个量称为微分,而把 $A$ 看成一个数即曲线 $M$ 在 $P$ 点的斜率,而正如我们在第三章中讲的,我们则把 $A$ 看成

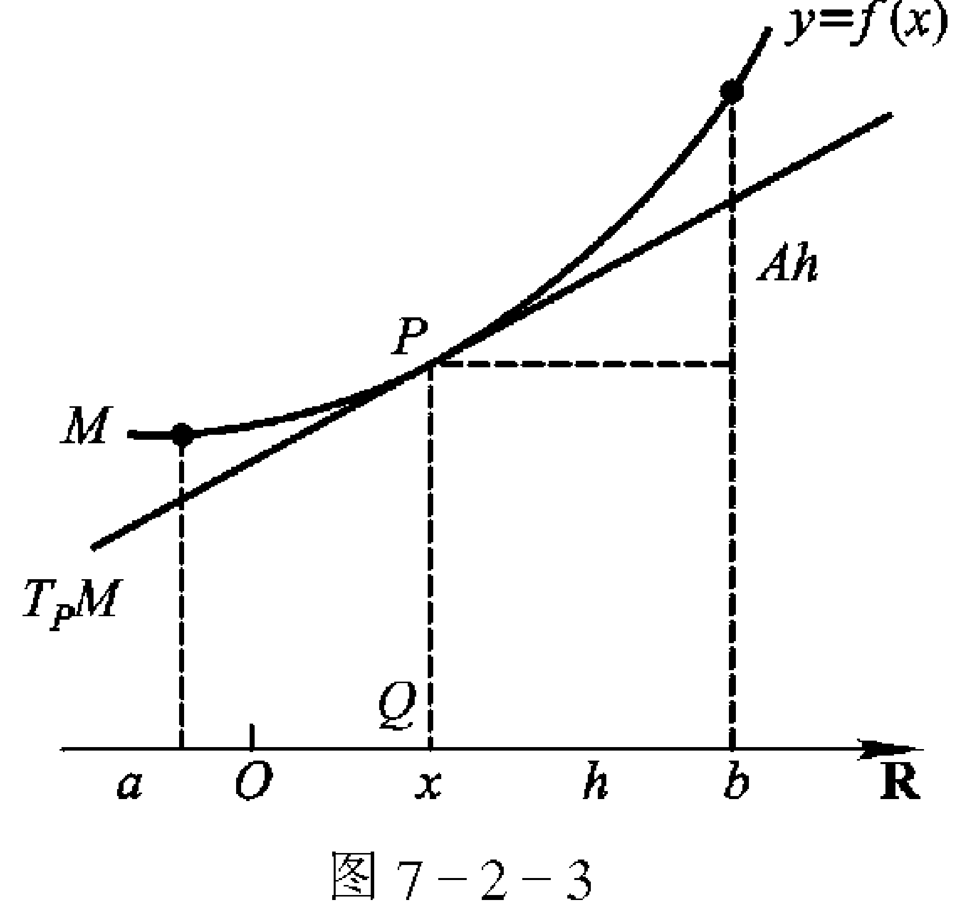


图 7-2-3

一个线性算子，它可以表示为一个一阶矩阵 $f'(x)$，而且就直接称此算子为微分. $h$ 则是切空间 $T_Q\mathbf{R}$ 中的向量. $Ah$ 则是微分这个线性算子作用到 $h$ 上以后得到的结果，它也是一向量. 不过，不论是通常的微积分教本还是本书，共同之处在于，$A$ 具有下述最根本的性质：

$$f(x+h)-f(x)=Ah+o(h). \tag{21}$$

现在我们怎样把它推广到两个微分流形 $M(\dim M=m)$ 到 $N(\dim N=n)$ 之间的可微映射？

$$f:M\to N,\quad P\in M\mapsto Q\in N$$

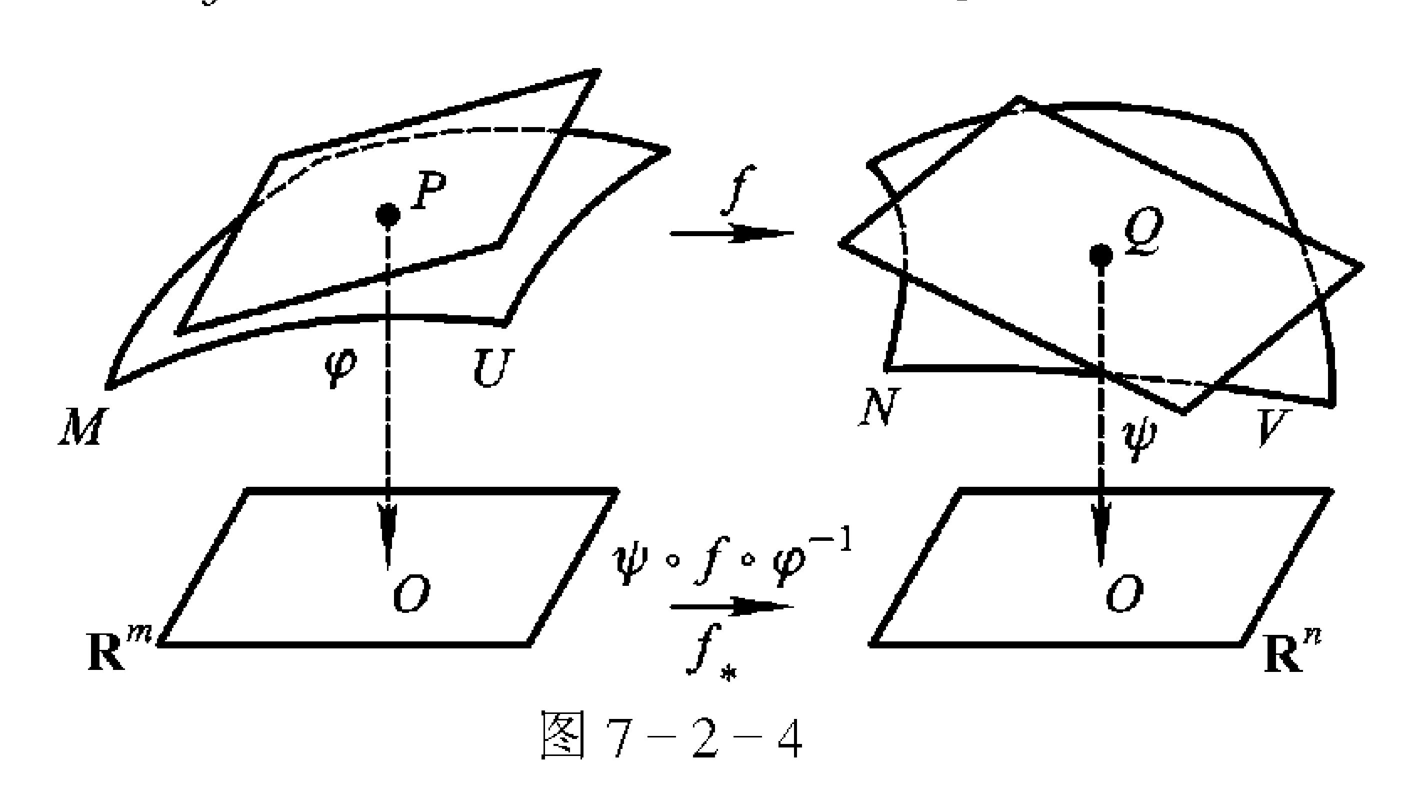



图 7－2－4

为了直观起见，我们仍设 $M$ 和 $N$ 各有一个维数充分高的 $\mathbf{R}^k$ 作为包含空间，并作出图 7－2－4：图上画出了 $M$ 和 $N$ 上两点的坐标邻域 $U$ 和 $V$ 以及坐标映射 $\varphi$ 和 $\psi$，并设 $\varphi(U)$ 和 $\psi(V)$ 中的局部坐标是 $x$ 与 $y$，而且 $P$，$Q$ 两点都映射到原点 $O$. 这与图 7－2－3 不同，那里 $P$ 点是映到 $x$ 处，不过我们认为 $x$ 是固定的. 图上在 $M$ 与 $N$ 的下方还画了 $\mathbf{R}^m$ 和 $\mathbf{R}^n$. 不过，它们中的每一个都起双重作用：一是作为微分流形与之同胚的那个空间：$\varphi(U)\subset\mathbf{R}^m$，$\psi(V)\subset\mathbf{R}^n$，它们是与 $U$，$V$ 同胚的. 二是上面我们已经说过，微分流形在每点上都有切空间而同构于 $\mathbf{R}^m$ 或 $\mathbf{R}^n$，这里的 $\mathbf{R}^m$ 和 $\mathbf{R}^n$ 就是 $T_PM$ 和 $T_QN$ 的同构像，它们与 $T_PM$，$T_QN$ 不仅是同胚而且线性同构. $\mathbf{R}^m$ 和 $\mathbf{R}^n$ 的这种双重作用的区别极为重要. 作为 $\varphi(U)$ 与 $\psi(V)$ 来看，对应于由 $M$ 到 $N$ 的映射 $f$，有一个映射，$\psi\circ f\circ\varphi^{-1}$，即 $f$ 的局部坐标表示. $\psi\circ f\circ\varphi^{-1}$ 是光滑的，其实这就是 $f$ 为光滑映射的定义. 作为 $T_PM$ 与 $T_QN$ 来看，其元是 $M$ 在 $P$ 点和 $N$ 在 $Q$ 点的切向量 $\xi$ 与 $\eta$，而当由 $M$ 到 $N$ 有映射 $f$（其坐标表示为 $\psi\circ f\circ\varphi^{-1}$）时，在 $\xi$ 与 $\eta$ 之间也应该有一个线性映射. 这就是图 7－2－3 中的 $A$ 的推广称为切映射. 正如图 7－1－3 中的 $A$ 称为 $f(x)$ 的微分，并在通常的微积分教本中记作 $\mathrm{d}f$：$A=\mathrm{d}f$（但实际上通常的微积分教本中的微分是 $Ah$）一样，这个我们认为应该存在的线性映射称为 $f$ 在 $P$ 点的切映射，也应该称为 $f$ 在 $P$ 点的微分，并记作 $\mathrm{d}f$（就是图 7－2－4 上的 $f_*$，这个记号将在下文中解释），同时也应该有与 (21) 的对应公式在.

现在我们就来寻找这个 $\mathrm{d}f$ 的相应的公式. 为此我们再回到切向量的定义，上文中我们用两个方法来定义切向量，其一是用光滑曲线 $c$ 的等价类 $[c]$，这里的 $c$ 要求当 $t=0$ 时经过 $P$ 点，等价关系则由同一等价类的两条曲线 $c_1$ 与 $c_2$ 必在 $P$ 点相切来定义. 但是我们所熟悉的相切性是欧氏空间 $\mathbf{R}^m$（我们先看 $T_PM$，$T_QN$ 的情况与此相仿）中的概念，所以 $c_1$ 与 $c_2$ 的相切需要先用坐标映射 $\varphi$ 将它们映到 $\varphi(U)\subset\mathbf{R}^m$ 中去，得到 $\mathbf{R}^m$ 中两条曲线

$$\varphi\circ c_i:t\to(x^1_{(i)}(t),\cdots,x^m_{(i)}(t)),\ i=1,2,$$

再按通常微积分教本的方法对 $t$ 在 $t=0$ 处求导得出

$$\xi_{(i)} = (\xi^1_{(i)}, \cdots, \xi^m_{(i)}) = (\dot{x}^1_{(i)}(0), \cdots, \dot{x}^m_{(i)}(0)), i = 1, 2,$$

是为切向量.如果 $\xi_{(1)} = \xi_{(2)}$(这一点下文还有说明),就说 $c_1$ 与 $c_2$ 在 $P$ 点相切,这时 $\xi_{(i)}$ 的公共值 $\xi \in \mathbf{R}^n$ 就称为$P$ 处的一个切向量,它是$[c]$的坐标表示.有了这一切再看什么是切映射就容易了,$f\circ c_1$ 与 $f\circ c_2$ 是 $N$ 上的两条光滑曲线,在坐标邻域中它们相应于 $\psi\circ f\circ c_i = (\psi\circ f\circ\varphi^{-1})\circ\varphi\circ c_i : t \mapsto (y^1_{(i)}(t), \cdots, y^n_{(i)}(t))$,而其对 $t$ 在 $t = 0$ 处的导数是 $(\eta^1_{(i)}, \cdots, \eta^n_{(i)}) = (\dot{y}^1_{(i)}(0), \cdots, \dot{y}^n_{(i)}(0))$,这里 $\psi\circ f\circ\varphi^{-1}$是 $\varphi(U)$到 $\psi(V)$的映射

$$y^k = y^k(x^1, \cdots, x^m), k = 1, 2, \cdots, n,$$

而

$$\dot{y}^k_{(i)}(0) = \frac{\partial y^k}{\partial x^j}(0)\dot{x}^j_{(i)}(0).$$

(请读者注意:一个重复出现的指标代表对该指标在适当范围中求和,所以这里出现的两个 $j$ 因为是$x$ 的指标,而 $x\in\mathbf{R}^m$,所以应从1 到 $m$ 求和;如果是对 $y\in\mathbf{R}^n$ 的指标求和,则要从1 到 $n$ 求和.)如果用向量来表示则得

$$\eta_{(i)} = A\xi_{(i)}, \quad i = 1, 2.$$

这里 $A$ 是雅可比矩阵$\dfrac{\partial(y^1, \cdots, y^n)}{\partial(x^1, \cdots, x^m)}(0)$.它确实是 $n\times m$ 矩阵而不是方阵.所以没有相应的行列式:

$$A = \begin{pmatrix} \dfrac{\partial y^1}{\partial x^1}(0) & \cdots & \dfrac{\partial y^1}{\partial x^m}(0) \\ & & \\ \dfrac{\partial y^n}{\partial x^1}(0) & \cdots & \dfrac{\partial y^n}{\partial x^m}(0) \end{pmatrix}. \tag{22}$$

因为 $\xi_1 = \xi_2 = \xi$,所以 $\eta_1 = \eta_2$,从而 $f\circ c_1 \simeq f\circ c_2$,其等价类为$[f\circ c]$,其坐标表示则是 $\eta_1$ 与 $\eta_2$ 的公共值 $\eta$,而上式成为 $\eta = A\xi$.在讲切空间的定义时我们已经说了,我们要把 $\xi\in\mathbf{R}^m$ 的线性构造移到$[c]$上去,使得$[c]$之集合也成一个线性空间,所以由于 $A$ 是$\xi\in\mathbf{R}^m$(注意现在的 $\mathbf{R}^m$ 是 $T_PM$ 而不是 $\varphi(U)$的包含空间)到 $\eta\in\mathbf{R}^n$ 的线性映射,所以由$[c]$到$[f\circ c]$之间有一线性映射,称为映射 $f$ 在$P$ 点的切映射,或称为 $f$ 在$P$ 之微分 $\mathrm{d}f_P$.$\mathrm{d}f_P$ 的坐标表示则是雅可比矩阵$A$((22)式).这是我们用切空间的第一种解释得到的切映射之定义的说明.

上文说,关于 $\xi_{(1)} = \xi_{(2)}$还有说明,这指的是,在通常微积分教本中,讲到两曲线相切时只需要其切线有相同斜率即可,所以只要 $\xi_{(1)} = C\xi_{(2)}$,$C$ 是一个非零实数,就应该说它们定义相同的切向量;或者说,不需要 $\xi_{(1)} = \xi_{(2)}$,只需要 $\xi_{(1)} /\!/ \xi_{(2)}$.而我们这里 $\xi = \dot{x}(0)$,其实是由速度的概念而来,所以,我们现在讲的切向量,其实是力学的速度向量,而不是几何向量.

我们现在来寻找公式(21)的类似物.在第三章中我们就说过,微分公式(21)中的 $x$ 与$h$ 性质不同:前者是底空间之元,后者则是切空间之元.二者本来无法相加.(21)中写出 $x + h$ 只是因为现在底空间是 $\mathbf{R}^n$ 的一个区域,切空间则是整个 $\mathbf{R}^n$,它们很偶然地可以相加.同样,$f(x + h) - f(x)$也只对这种特殊情况有意义.在 $f$ 表示微分流形$M$ 到另一微分流形$N$ 的可微

映射时，应怎样寻找(21)的类似物呢？为此，我们要用 $f$ 的局部坐标表示，即 $\psi\circ f\circ\varphi^{-1}$ 并将它和前面一样写成

$$y = y(x) \quad 或 \quad y^i = y^i(x^1, \cdots, x^m), i = 1, 2, \cdots, n. \tag{23}$$

我们还设 $P$ 和 $Q$ 分别即是 $x=0$ 与 $y=0$，于是在 $x=0$ 处把(23)用泰勒公式写成

$$y^i = \frac{\partial y^i(0)}{\partial x^j}x^j + 高阶项.$$

并取它的线性部分(注意，我们不再说“略去高阶项”的话了，第三章第二节作了详细说明)：

$$\eta = A\xi, \qquad A = \frac{\partial(y^1, \cdots, y^n)}{\partial(x^1, \cdots, x^m)}(0).$$

这里我们用 $\eta$ 和 $\xi$ 取代了 $x$ 和 $y$. 事实上，$x$ 与 $y$ 是底空间中之元的局部坐标，$f$ 本来就是底空间之间的映射；$\eta$ 和 $\xi$ 则是切向量——切空间之元的局部坐标，而切映射正是切空间之间的线性映射. 至此我们更看到切映射概念确是函数的微分的概念之推广. 而且用局部坐标表示后更表明它们的确是一回事. 我们不说略去了 $o(h)$，而说成是取泰勒公式的 1-节(1-jet)来代替 $f$，而 1-节的形状是与局部坐标的选取无关的(此即一阶微分的形式不变性). “略去高阶无穷小”现在代之以“只考察 1-节”. 这样绕过了无法定义 $f(x+h)-f(x)$ 的困难. 总之，切映射就是微分概念的推广，所以我们采用了同样的记号 $\mathrm{d}f$.

但是，切向量还可以用一阶微分算子作定义. 按这种定义切映射又应如何理解呢？它能告诉我们什么新的联系、新的概念吗？为此，我们暂时把有待我们去探讨的切映射记作 $\tilde{\mathrm{d}}f$，并且打算说明它就是 $\mathrm{d}f$. 于是设有 $X_P = \xi^i\dfrac{\partial}{\partial x^i} \in T_PM$. 这里 $(\xi^1, \cdots, \xi^m) = (\dot{x}^1(0), \cdots, \dot{x}^m(0))$ 是它的局部坐标表示. 于是，$\tilde{\mathrm{d}}f\cdot X_P$ 应是 $N$ 在 $Q$ 点的切向量，也不妨设它在 $N$ 上的局部坐标表示是 $Y_Q = \eta^k\dfrac{\partial}{\partial y^k}, k=1,2,\cdots,n$. 于是我们问，应该怎样决定 $\eta=(\eta^1, \cdots, \eta^n)$？一个微分算子是通过它在函数上的作用来表现的，所以对任意的定义在 $f(M)\subset N$ 上的函数 $F$(不妨用局部坐标 $F(y)$ 表示)，$Y_QF = \eta^k\dfrac{\partial F}{\partial y^k}$ 是有定义的. 但是 $F$ 也可以认为是定义在 $M$ 上的，即先把 $M$ 上之点($x$ 是它的局部坐标)通过映射 $f$ 映到 $N$ 上，而原来在 $N$ 上的函数 $F(y)$ 现在变成 $F(y(x))$，被拉回到 $M$ 上. 一个十分自然的想法是这样来定义 $\tilde{\mathrm{d}}f\cdot X_P = Y_Q$：使 $X_P$ 作用到 $F(y(x))$ 上恰好就是 $Y_Q$ 作用到 $F(y)$ 上. 亦即在 $x=0$ 及其像 $y=0$ 处

$$\eta^k\frac{\partial F(y)}{\partial y^k}\bigg|_{y=0} = \left(\xi^i\frac{\partial}{\partial x^i}\right)F(y(x))\bigg|_{x=0}$$

$$= \xi^i\frac{\partial F}{\partial y^j}\frac{\partial y^k}{\partial x^i}\bigg|_{y=y(x),x=0}.$$

因为此式应该对任意函数 $F(y)$ 成立，依次令 $F=y^1, \cdots, y^n$，即有

$$\eta^k = \xi^i\frac{\partial y^k(0)}{\partial x^i}.$$

但是这正是用第一种方法得出的 $\mathrm{d}f$ 的局部坐标表示 $\eta = A\xi$. 所以，切映射的这两种定义是完全一致的. 我们也就没有必要再用另一个记号 $\tilde{\mathrm{d}}f$，而都用 $\mathrm{d}f$. 第二种讲法实际上引入了一种对偶性. 因为 $\mathrm{d}f_P$ 把 $M$ 上的切向量 $X_P$ 推前到 $N$ 上，成为 $N$ 上的切向量. $F(y(x))$ 或记为 $F(f)$

$=F\circ f$则把定义在$N$上的函数$F$拉回到$M$上成$F\circ f$.第二种定义就是说:或者把$X_P$推前,或者把$F$拉回,二者相同.这是一种深刻的对偶性,见后文.

以上讲的既是切映射的定义及其解释,也证明了它的一些性质.为明确起见,我们把它归结为下面的

**定义和定理 4** 设 M 和 $N$ 分别是 $m$ 维与 $n$ 维微分流形. $f: M\to N, P\mapsto Q$ 是一个可微映射.则

(1) $M$ 中经过$P$ 的光滑曲线$c$之集合中可定义一等价类$[c]$如上,必有线性映射 $\mathrm{d}f_P: T_PM\to T_QN$ 映$[c]$为 $T_QN$ 中的元. $\mathrm{d}f_P$ 称为$f$在$P$之切映射,也称为 $f$ 在$P$ 之微分.

(2) $\mathrm{d}f_P$ 是映射$f$ 的线性部分,而且其局部坐标表示是

$$\eta = \mathrm{d}f_P\cdot\xi = A\xi,\xi\in T_PM,\eta\in T_QN, \tag{24}$$

$A$ 是雅可比矩阵: $A=\left(\dfrac{\partial y^i(0)}{\partial x^j}\right)$, $\mathrm{d}f_P\cdot\xi$ 是 $y=y(x)$在 $x=0$ 处的泰勒展开式的 1 - 节.

(3) 若用一阶微分算子 $X_P$ 表示 $T_PM$ 中之元,则 $\mathrm{d}f_P\cdot X_P$ 是 $T_QN$ 中的下述微分算子:

$$\mathrm{d}f_P\cdot X_P(F) = X_P(F\circ f). \tag{25}$$

**定理 5** 切映射 $\mathrm{d}f$ 具有以下性质:

(1) 链式法则 若 $f_1: M_1\to M_2, P_1\mapsto P_2, f_2: M_2\to M_3, P_2\mapsto P_3$ 是两个可微映射,则

$$\mathrm{d}(f_2\circ f_1)(P_1) = \mathrm{d}f_2(P_2)\cdot \mathrm{d}f_1(P_1).$$

(2) 莱布尼茨性质 若 $F,G$ 是$Q\in N$ 附近的光滑函数,$f: M\to N$ 是一可微映射,$f(P)=Q$,则

$$(\mathrm{d}f_P\cdot X_P)(FG)(Q) = F(Q)(\mathrm{d}f_P\cdot X_P)G(Q) + G(Q)(\mathrm{d}f_P\cdot X_P)F(Q).$$

这些性质由 $\mathrm{d}f$ 之定义都容易直接证明,所以这里都省略了.

现在回到古典的微分概念.一个定义在微分流形 $M$ 上的可微函数可以看作是可微映射 $f: M\to\mathbf{R}^1$,即 $N=\mathbf{R}^1$ 的特例.这时 $n=\dim N=1$,且 $T_QN=\mathbf{R}^1$.于是设有 $M$ 上的切向量 $X_P$, $\mathrm{d}f_P\cdot X_P$应该是 $N=\mathbf{R}^1$ 上的切向量.由于 $T\mathbf{R}^1=\mathbf{R}^1$,其上有一个独立参数 $t$,所以 $T\mathbf{R}^1=\{\lambda\dfrac{\mathrm{d}}{\mathrm{d}t}\}$, $\lambda$ 是实数.于是必存在一个实数 $\lambda$ 使

$$\mathrm{d}f_P\cdot X_P = \lambda\frac{\mathrm{d}}{\mathrm{d}t}.$$

我们选一个定义在 $N$ 上的函数$F\equiv t$.用上面的微分算子作用于它,有

$$(\mathrm{d}f_P\cdot X_P)(t) = \lambda.$$

但由定义, $t$"拉回"到 $M$ 上,就成为 $f_P$.因此有

$$(\mathrm{d}f_P\cdot X_P)(t) = X_P(f).$$

上面我们说了,切向量可以有两个方法定义,一是作为一个微分算子 $X_P=\xi^i\dfrac{\partial}{\partial x^i}$,二是定义为它的系数所成的向量 $\xi=(\xi^1,\cdots,\xi^m)=\dot{x}$.在我们这里,前者是 $\lambda\dfrac{\mathrm{d}}{\mathrm{d}t}$,后者则是 $\lambda$,可以认为它们都是一样的.在这个意义下,我们有

**定理 6** 若 $f$ 是$M$ 上的光滑函数,$P\in U\subset M$,则对一切 $X_P\in T_PM$,

$$df_P \cdot X_P = X_P(f). \tag{26}$$

尽管从表面上看上式左方是一个算子，右方则是一个实数，二者怎能相等？有了这样的说明后，即有

**推论 7** 假设同上，有

$$df_P = \frac{\partial f(0)}{\partial x^i} dx_P^i. \tag{27}$$

**证** 在(26)中令 $f$ 为第 $j$ 个坐标函数 $x^j$，$X = \dfrac{\partial}{\partial x^i}$，即有

$$dx^j \left(\frac{\partial}{\partial x^i}\right) = \delta_j^i. \tag{28}$$

对任意的 $X_P = \xi^i \dfrac{\partial}{\partial x^i}$，设 $P$ 即 $x = 0$，

$$df_P \cdot X_P = X_P(f) = \xi^i \frac{\partial f(0)}{\partial x^i} = \delta_j^i \xi^j \frac{\partial f(0)}{\partial x^i} = \frac{\partial f(0)}{\partial x^i} dx_P^i \xi^j \frac{\partial}{\partial x^j} = \frac{\partial f(0)}{\partial x^i} = dx_P^i \cdot X_P.$$

此即(27)式.

这样一方面我们得到了与古典的微分公式形状完全相同的结果；另一方面又大大发展了第三章所讲的微分学的基本思想，即“略去高阶无穷小”. 但是我们所得又远不止于此. 我们把切向量变成了一个一阶微分算子，而切映射，作为函数微分的推广并不只是把 $X_P$ 映为另一个微分算子例如 $Y_Q$，而且如定理 6 指出的那样，在经过了一定的解释以后，对任意的 $X_P \in T_PM$ 给出一个实数值 $X_P(f)$. 这就是(26)式的含意. $X_P(f)$ 对 $X_P$ 当然是线性的：$[c_1 X_P^{(1)} + c_2 X_P^{(2)}](f) = (df_P)(c_1 X_P^{(1)} + c_2 X_P^{(2)}) = c_1 X_P^{(1)}(f) + c_2 X_P^{(2)}(f)$，所以 $df$ 是 $T_PM$ 上的线性泛函，而且因为 $\dim T_PM = \dim M = m$ 是有限的，所以还是 $T_PM$ 上的连续线性泛函. 这样，$df_P$ 就成了 $T_PM$ 的对偶空间 $T_P^*M$ 中的元. 上面我们说切映射的研究导致了一种深刻的对偶性的出现就是指此.

$T_P^*M$ 之元称为余切向量，其空间 $T_P^*M$ 称为余切空间. 余切空间作为 $T_PM$ 之对偶空间其维数与 $T_PM$ 相同，即同为 $\dim M = m$，现在要问，$df_P$ 是否已概括了全部余切向量？或者说，余切向量是否全由某个光滑函数 $f$ 之微分生成？当然不是如此，这是一个很深刻的问题，我们将在 §4 中讨论它. 在目前我们只需指出一点，我们现在讨论余切向量是限制固定一个 $P$ 点的，而当 $P$ 在 $M$ 上变时，这个余切向量也会变动，这时是否所有这些余切向量都是由同一个函数之微分生成是很不确定的. 在古典的微积分教本中我们已见到这种问题. 例如，设有 $x dy - y dx$，这里 $(x, y) \neq (0,0)$，当 $(x, y)$ 变动时能不能找到一个函数 $\varphi(x, y)$ 使 $x dy - y dx = d\varphi$？当然不行，因为如果有这样的光滑函数 $\varphi$ 存在，当有 $x = \dfrac{\partial \varphi}{\partial y}$，$-y = \dfrac{\partial \varphi}{\partial x}$ 而 $\dfrac{\partial^2 \varphi}{\partial y \partial x} = -1$，$\dfrac{\partial^2 \varphi}{\partial x \partial y} = 1$ 二者不相等. 但是若用 $\dfrac{1}{x^2 + y^2}$ 去乘此式，得到 $\dfrac{x dy - y dx}{x^2 + y^2} = d\left(\arctan \dfrac{y}{x}\right)$. 前面要求 $(x, y)$ 不是原点原因在此. 我们不去讨论当 $P$ 变动时怎样求出各点的余切空间中的全部元素，而要问，当 $P$ 固定时，怎样求 $T_P^*M$ 的基底. 由于 $T_P^*M$ 维数为 $m$，所以我们想去求出 $T_P^*M$ 的 $m$ 个线性无关的元. (27)告诉我们 $df_P$ 是 $\{dx_P^i\}$，$i = 1, 2, \cdots, m$，的线性组合，所以只要得证 $\{dx_P^i\}$ 是线性无关的则它们必是 $T_P^*M$ 的一个基底.

因此，设有实数 $\lambda_1, \cdots, \lambda_m$ 使

$$\lambda_j \mathrm{d}x_P^j = 0,$$

把双方作用到$\dfrac{\partial}{\partial x^i}$上，由(28)即得

$$\lambda_i = 0, i = 1, 2, \cdots, m.$$

因此$\{\mathrm{d}x_P^i\}$，$i=1,2,\cdots,m$，是 $T_P^*M$ 之基底，概括以上所述，我们有

**定义与定理 8** $T_PM$ 之对偶空间 $T_P^*M$ 称为 $M$ 在 $P$ 点的余切空间，若在 $M$ 上某个坐标邻域中取局部坐标$\{x^i\}$，$i=1,2,\cdots,m$，则$\{\mathrm{d}x_P^i\}$是 $T_P^*M$ 的一个基底，从而 $T_P^*M$ 之一切元，即 $M$ 在 $P$ 点的一切余切向量必可写为

$$\omega_P = \lambda_i(P)\mathrm{d}x_P^i. \tag{29}$$

**注 1** 当 $P$ 变动时 $\omega_P$ 不一定是某个函数之微分，即若不附加其它条件，$\lambda_i(P)$不会是某个光滑函数 $\varphi$ 之全部偏导数. 所以(29)称为微分形式而不是微分.

**注 2** 在局部坐标 $x$ 中，$\left\{\dfrac{\partial}{\partial x^i}\right\}$与$\{\mathrm{d}x^i\}$是 $T_PM$ 与 $T_P^*M$ 之一对对偶基底.

**注 3** 至此我们应该提一下切(余切)向量究竟是协变的还是逆变的? $T_PM$ 的一个基底是$\left\{\dfrac{\partial}{\partial x^1}, \cdots, \dfrac{\partial}{\partial x^m}\right\}$，按我们前面的讲法是说$\dfrac{\partial}{\partial x^i}$是协变的. 但是前面讲的是$\dfrac{\partial f}{\partial x^i}$，它是梯度向量 **grad** $f$的分量，即以$\{e_1, \cdots, e_m\}$为基底时的坐标. 现在的$\dfrac{\partial}{\partial x^i}$本身就是一个向量，它是标架$\left\{\dfrac{\partial}{\partial x^1}, \cdots, \dfrac{\partial}{\partial x^m}\right\}$的第 $i$ 个元. 它不是某一个向量的坐标，它自己却有坐标. 例如$\dfrac{\partial}{\partial x^1} \cong (1, 0, \cdots, 0)$. $T_PM$ 的任一元 $X_P = \xi^i \dfrac{\partial}{\partial x^i}$，$\xi^i$ 才是其相对于标架$\left\{\dfrac{\partial}{\partial x^1}, \cdots, \dfrac{\partial}{\partial x^m}\right\}$的坐标，$\dfrac{\partial}{\partial x^i}$是协变的，其坐标则一定是逆变的. 这样，$X_P$ 才是与坐标无关的，无所谓协变与逆变. 同样，$T_P^*M$ 中的任一元 $\omega = \lambda_i \mathrm{d}x^i$，$\{\mathrm{d}x^1, \cdots, \mathrm{d}x^m\}$是标架，每一个 $\mathrm{d}x^i$ 都是一个向量，而且是逆变的. 这样，为了使 $\omega$ 成为与坐标无关的，$\omega$ 对于上述基底的坐标$\{\lambda_1, \cdots, \lambda_m\}$又一定是协变的.

最后引入一个记号，我们记 $\mathrm{d}f_P$ 为 $f_*$，记 $F \circ f = f^* F$，$f_*$ 是一个"推前"(push forward)，因为 $f$ 把$M$ 映成$N$，$f_*$ 则把 $T_PM$ 映成$T_QN$，"方向"是一致的. $f^*$ 则是一个"拉回"(pull back)，因为若 $F$ 定义在$N$ 上，则 $F \circ f$ 定义在$M$ 上，所以 $f^* F$ 把原来定义于$N$ 上的"对象"(现在是一个函数)拉回到 $M$ 上来了. 这样一来(25)式就可以写成

$$\langle f_* X_P, F \rangle = \langle X_P, f^* F \rangle.$$

它是说，或者把 $M$ 上的切向量$X$ 推前到$N$ 上成为$f_* X \in T_QN$，然后把它作用到定义在 $N$ 上的函数$F$ 上去；或者向量 $X_P$ 不动而把定义在$N$ 上的$F$ 用$f^*$ 拉回到 $M$ 上，再与 $X_P$ 作用. (25)表明，这两种方法结果是一致的.

用现在的记号就发现 $f_*$ 与 $f^*$ 的关系恰好是互相对偶的. 我们又时常把 $f_*$ 写成 $T_Pf$，于是切映射就有了几种记法：$\mathrm{d}f(P)$，$f_*(P)$或 $T_Pf$ 或$(Tf)(P)$.

**3. 切丛与余切丛** 我们多次讲过，在函数的微分的定义中出现的 $f(x+h)$，$x$ 与$h$ 必须从

概念上加以区别.现在可以解释为什么要这样做了.设有微分流形 $M$,其维数为 $n$,令 $U\subset M$ 是其上的一个区域.对于每一点 $P\in U$,都可以作其切空间 $T_PM$.把它们并起来,并记所得为 $T_UM=\bigcup_{P\in U}T_PM$,称为 $U$ 上的切丛(特别当 $U=M$ 时,我们就记之为 $TM$).这个集合在物理,力学和几何学中是很有用的,它的元素即"点"为 $(x,v)$ 其中 $x\in U$,$v\in T_xM$,如果用局部坐标表示即为 $(x^1,\cdots,x^n;v^1,\cdots,v^n)$,这里 $x$ 和 $v$ 就是分开来的.在古典的微分学中所讲的 $f(x+h)$,$x$ 与这里讲的 $U$ 中之点相应,$h$ 则与这里讲的 $v$ 相应.只是因为在古典的微分学中,$x\in U=(a,b)$,$h\in\mathbf{R}$(这里 $\mathbf{R}$ 作为一个集合即前面的实数域,作为一个线性空间即 $\mathbf{R}^1$,下面对它们不作区别都记为 $\mathbf{R}$)而确实可以作加法.这样 $x+h$ 其实是表示以 $x$ 点为原点作向量 $h$,亦即在 $x$ 点附加上曲线 $y=f(x)$ 过 $(x,f(x))$ 的切线,这样就得到曲线 $y=f(x)$ 之切丛.在讲微分概念时,要固定 $x$ 是因为限于在某点的切空间中讨论问题,用不到切从.现在有了微分流形的切丛概念以后才发现 $(x,v)$ 中 $x$ 与 $v$ 本来就是泾渭分明的:$x\in M$,对它可以作微分流形理论所允许的各种操作,但就是不能作加法,因为 $M$ 一般没有线性结构;反之,$v\in T_PM\simeq\mathbf{R}^n$,在同一点 $x$ 的各个不同 $v$ 之间则允许作线性运算,但对不同点的 $v$ 例如对 $(x_1,v_1)$ 和 $(x_2,v_2)$,当 $x_1\neq x_2$ 时 $v_1+v_2$ 就是没有意义的,更不说 $x+v$ 了.以上的做法就是在 $M$ 的每一点上附上一个线性空间 $T_PM$;当然也可以附加 $T_P^*M$ 而得到余切丛 $T_U^*M$ 或 $T^*M$;甚至可以附上其它的什么,例如附上一条直线得到线丛,附上一个球得到球丛,诸如此类都是一般所谓纤维丛之特例.这在物理学与几何学中是极大的进步,但已远远超出了本书所可能涉及的范围.我们只能简单地提一下.这类几何对象已经远远地超过我们的直觉.当然有一些极简单的情况,例如一条空间曲线 $M$(注意 $M$ 是在 $\mathbf{R}^3$ 中扭曲的,不是平放在纸面上的平面曲线),其在 $P$ 点的切线一部分在纸面上方,另一部分在纸面下侧,本是看不见的,合起来成了一个如镰刀刃口那样的曲面,这个 2 维曲面就是切丛 $TM$.图 7-2-5 上用虚线画的是切线被曲面前一叶遮住而看不见的部分.$M$ 称为这个曲面的脊线(edge of regression).这种几何图形在微分几何的曲线理论中会出现,在奇点理论中也会出现,是切丛几何形象的一个好例子.但除此以外还有什么"看得见摸得着"的例子?至于余切丛,因为余切空间就已经没法用几何表示了,更不必说余切丛了.然而几何直观又确实是很有用的,所以我们不妨把这些切空间都画成竖直的直线,于是如图 7-2-5 的下方,$TM$ 反而成了一个"柱面"$M\times\mathbf{R}$,这当然"很不像"了,但却比"看得见摸得着"的图形更准确.因为例如 $S^1$ 的切丛,如果真要画出来,就是 $\mathbf{R}^2$ 上挖出了单位圆盘,但却是由两叶合成的,因为经过圆外一点可以向此圆作两条切线,所以这一点既在 $T_{P_1}S^1$ 上,又在 $T_{P_2}S^1$ 上.反而圆周上的点只在一条切线上,所以 $TS^1$ 好像是把那个镰刀刃口压平了的平面区域.就 $TS^1$ 应为 2 维流形这点来说,这个图形倒很真实,但是说它与柱面 $S^1\times\mathbf{R}$ 同构(这是真的),却是看不出来.画成图 7-2-5 的下图读者可能会感到,$T_PM$ 与 $M$ 相切于 $P$ 点没有画出来,但是再一想,把微分流形画在 $\mathbf{R}^2$ 或 $\mathbf{R}^3$ 中本来就不一定办得到,有图 7-2-5 下图反倒很有利.所以一般书上讲切丛、余切丛乃至一般的纤维丛总是用它以示意.

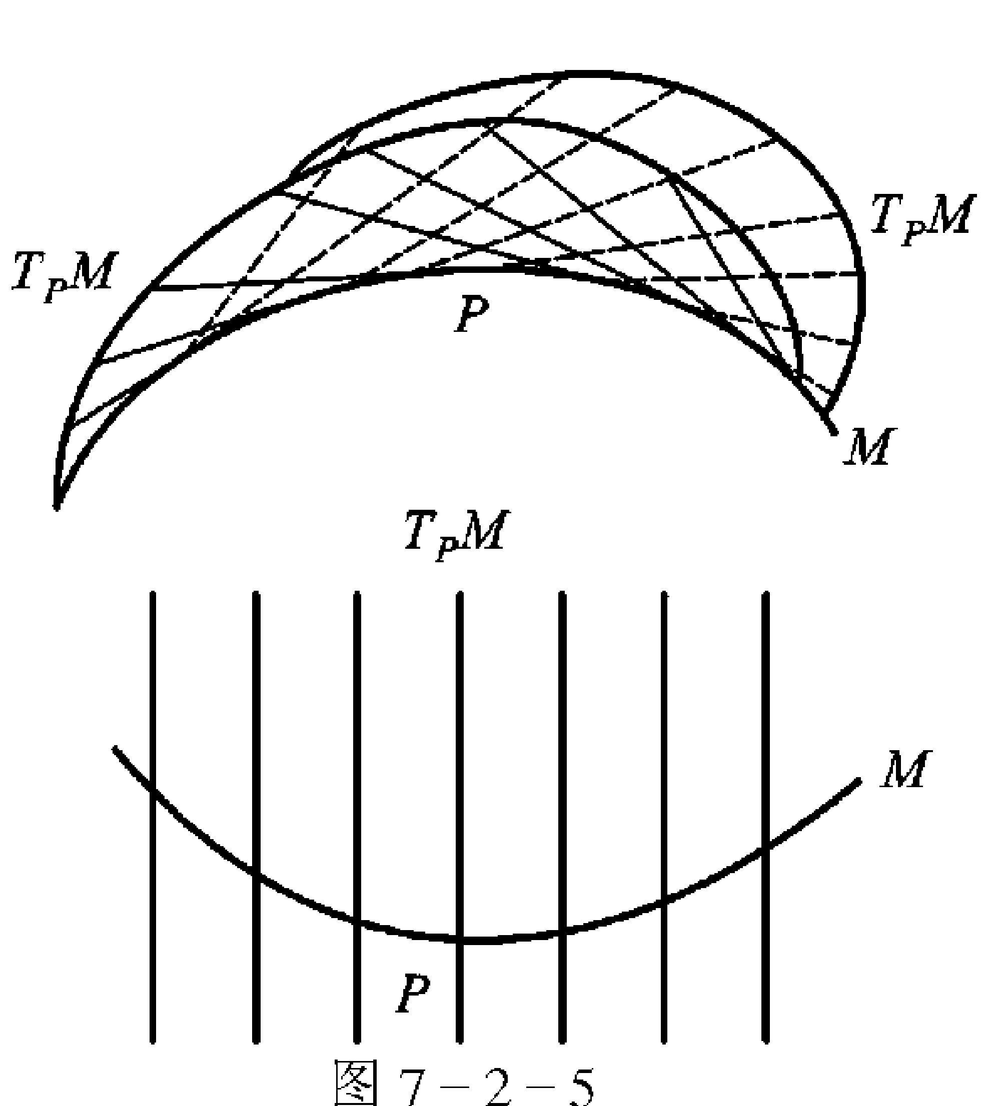


图 7-2-5

上面我们说到切丛与余切丛的概念在力学上是很有用的,

实际上可以说正是力学和物理学问题启发了这些概念.例如讨论一个自由度为 $n$ 的力学系,为了刻画它的状况,我们需要知道这系统在时刻 $t$ 的位置,这是由它的 $n$ 个广义坐标$\{(q^1(t),\cdots,q^n(t))\}$来刻画的.这个集合通常是一个 $n$ 维流形$M$ 而不一定是 $\mathbf{R}^n$ 的某个区域.这是因为这个力学系统会受到某些约束,例如被限制在某一曲面上运动.为了刻画这个系统,还需要了解它的广义速度$\{(\dot{q}^1(t),\cdots,\dot{q}^n(t))\}$.如果固定一点 $q$,则一切可能的速度成为 $T_qM$,而如果 $q$ 变化,则应该用一个 $2n$ 维空间中的点$(q^1,\cdots,q^n,\dot{q}^1,\cdots,\dot{q}^n)$来刻画它.这些点显然在 $M$ 的切丛 $TM$ 内(但不一定是整个 $TM$,因为速度也可能受到某些约束,例如某些守恒律的限制),$M$ 称为这个力学系统的构形空间(configuration space).第一节中我们还说过,描述一个力学系统有时应该用广义动量$(p_1(t),\cdots,p_n(t))$.广义动量是一个协变向量而广义速度$(\dot{q}^1(t),\cdots,\dot{q}^n(t))$是一个逆变向量.协变向量与逆变向量的内积是一个标量.因此 $p(t)$成为 $T_PM$ 上的线性泛函.所以,广义动量是一个余切向量,而$(q^1,\cdots,q^n;p_1,\cdots,p_n)\in T^*M$.所以,如果我们利用广义动量来讨论力学系统,余切丛就成了合适的框架.这时,$M$ 仍称为构形空间,余切丛 $T^*M$ 则称为相空间(phase space).(切丛一般不应称为相空间.)经典力学是这样,量子力学也是这样.不过它一般需要无限维空间而更加复杂了.

可见,研究切丛与余切丛的几何结构,不只在数学上而且在物理和力学上都有重要的意义.下面我们的目标是证明,它们都是一类很特殊的微分流形.由于余切丛的讨论与切丛的讨论是平行的,所以我们只讨论切丛.

首先,切从(余切丛)是切空间(余切空间)之并:

$$TM = \bigcup_{P\in M} T_PM,\quad T^*M = \bigcup_{P\in M} T_P^*M. \tag{30}$$

我们称 $TM(T^*M)$为全空间 $E$,$M$ 为底空间(base space).这些丛都有一个投影算子 $\pi$:

$$\pi: E\to M,\qquad (x,v)\in T_PM\mapsto x\in M.$$

而底空间 $M$ 上一点$P$ 在投影算子下的原像$\pi^{-1}(P)=T_PM(T_P^*M)$必同构于 $\mathbf{R}^n$,称为 $E$ 在$P$ 点处的纤维(fiber).为了证明 $E$ 是微分流形,就应构造出它的坐标邻域 $V_\alpha$,这个坐标邻域中的坐标映射 $\Psi_\alpha$,以及证明迁移函数 $\Psi_\beta\circ\Psi_\alpha^{-1}$ 是一个微分同胚.所有这一切我们都将从 $M$ 的坐标邻域$U_\alpha$,坐标映射 $\varphi_\alpha$ 以及迁移函数$\varphi_\beta\circ\varphi_\alpha^{-1}$ 出发来完成.

我们定义 $E$ 的坐标邻域$V_\alpha$ 与坐标映射$\Psi_\alpha$ 如下:

(1) $V_\alpha=\pi^{-1}(U_\alpha)$,$E$ 上的拓扑应该这样定义,使得 $\pi:E\to M$ 是连续映射,这样 $V_\alpha$ 作为$M$ 中开集 $U_\alpha$ 在连续映射下的原像故为$E$ 中的开集,且

$$\bigcup_\alpha V_\alpha = E. \tag{31}$$

(2) 我们要求 $V_\alpha$ 与 $U_\alpha\times\mathbf{R}^n$ 为同胚.因此存在坐标映射 $\Psi_\alpha: V_\alpha\to\varphi_\alpha(U_\alpha)\times\mathbf{R}^n\subset\mathbf{R}^n\times\mathbf{R}^n\simeq\mathbf{R}^{2n}$,而 $V_\alpha$ 中之一点$Q$ 有局部坐标$(x^1,\cdots,x^n;\xi^1,\cdots,\xi^n)$.实际上,由于 $E=\bigcup_P T_PM$,故 $Q=(x,X)$,$x$ 是$M$ 中某一点,其在 $\varphi_\alpha$ 中之局部坐标为$(x^1,\cdots,x^n)$,而 $X$ 是$x$ 点处的一个切向量 $X=\xi^i\dfrac{\partial}{\partial x^i}$.

(3) $V_\alpha\cap V_\beta$ 中的迁移函数 $\Psi_{\beta\alpha}=\Psi_\beta\circ\Psi_\alpha^{-1}$ 是这样定义的:对于 $V_\alpha\cap V_\beta$ 中同一点,有两个

不同的局部坐标$(x_\alpha,\xi_\alpha)$与$(x_\beta,\xi_\beta)$. 设这一点在底空间$M$上之投影为$P$,则$x_\alpha$与$x_\beta$分别是$P$在$(U_\alpha,\varphi_\alpha)$与$(U_\beta,\varphi_\beta)$下的局部坐标,而$x_\beta=\varphi_\beta\circ\varphi_\alpha^{-1}(x_\alpha)$,$\xi_\alpha=\mathrm{d}\varphi_\alpha(P)X_P$,$\xi_\beta=\mathrm{d}\varphi_\beta(P)X_P$,所以$\xi_\beta=(\mathrm{d}\varphi_\beta\circ\mathrm{d}\varphi_\alpha^{-1})(P)\xi_\alpha=\mathrm{d}\varphi_{\beta\alpha}(P)\xi_\alpha$. $\mathrm{d}\varphi_{\beta\alpha}(P)$当$P$变动时显然对$x\in U_\alpha\cap U_\beta$是$C^\infty$的. 于是我们得知

$$\Psi_{\beta\alpha}=(\varphi_{\beta\alpha},\mathrm{d}\varphi_{\beta\alpha})$$

是$V_\alpha$到$V_\beta$中的$C^\infty$映射. 在底空间上它就是$M$的迁移函数,在纤维上则是一个$C^\infty$依赖于$x$的线性同构.

**定义与定理 9**　设$M$为一个微分流形,维数为$n$,则$\bigcup\limits_P T_PM$ $(\bigcup\limits_P T_P^*M)$有如上的微分流形构造,维数为$2n$,称为$M$之切丛(余切丛).

以上构造$TM$与$T^*M$时的第二点:$\Psi_\alpha: V_\alpha\to\varphi_\alpha(U_\alpha)\times\mathbf{R}^n$最值得注意. 当然,在每点$P$处都有切空间$T_PM$(余切空间$T_P^*M$)与$\mathbf{R}^n$同构,故当$P$点变化,各点的$T_PM$互相亦同构,因而若给出一个固定的$\mathbf{R}^n$,这些$T_PM\simeq\mathbf{R}^n$,而且这个同构关系在局部坐标中可以用一非奇异矩阵$A(x)$表示. 切丛的定义要求$A(x)$之元是$x$之光滑函数,否则就得不出迁移函数之纤维部分是光滑的$\mathrm{d}\varphi_{\beta\alpha}$. 这样,$\mathbf{R}^n$适用于$P\in U_\alpha$时的所有纤维$T_PM$均与$\mathbf{R}^n$同构而与$x$无关. 这样$\varphi_\alpha(U_\alpha)\times\mathbf{R}^n$才有乘积形式. 这个情况习惯地说是:切丛必是局部平凡的(trivial locally). 那么可以问,一个微分流形$M$的切丛可否是整体平凡的(trivial globally),即整体地同胚于$M\times\mathbf{R}^n$(整体平凡的又称为可平行化(parallelizable)的)? 不可能,下面是一个简单而重要的例子.

$S^1$的切丛,上面我们说了,直观地可以认为是压平了的镰刀曲面. 但是若将$S^1$之每一点的切线(不妨规定按反时针方向取),都向上竖起$\frac{\pi}{2}$,我们把这个操作称为"竖起变换",很容易看到它是同胚,因而$S^1$的切丛必然同胚于"竖起变换"的结果,即一个圆柱: $TS^1\cong S^1\times\mathbf{R}$,因此$TS^1$是整体平凡的. 但是如果稍作一点变动就发现情况大相径庭了. 我们在$S^1$上取一点$P$,作一个竖起的切线(见图 7-2-5 下图),为简单起见在其上取一个有限长线段且以$P$为中心. 现在让$P$依反时针方向绕$S^1$运动而且一面运动一面向外侧倾倒,运动到角度为$\theta$时,要求向外倾倒$\frac{\theta}{2}$. (其实在舞蹈表演中常看见这样的动作.)这样当$P$点回到原处后,这个线段(这个演员)就会恰好翻过来了,头朝地脚朝天,成为一个默比乌斯带. 所以默比乌斯带可以看成一个线丛. 但是它决不可能与$TS^1$同胚,因此不会是整体平凡的. 例如,$TS^1$"拦腰切断"将会成为两个分离的圆柱,而把默比乌斯带沿$S^1$剪开——也就是"拦腰切断"——将成为两个套在一起的默比乌斯带. 读者自己可以试一下,可是,这样的区别不是一个小游戏,在物理问题中它可能反映一些根本的性质上的区别. 由此可见,切丛和余切丛内容非常丰富.

最后我们再介绍一个向量丛(包括切丛与余切丛)的光滑切口(cross-section)的概念. 这就是由$M$到$E$中的一个光滑映射$\sigma$,而且$\pi\circ\sigma=\mathrm{id}: M\to M$. 对于切丛,例如在一个坐标邻域$U$内,其光滑切口必可表为$X(x)=\xi^i(x)\frac{\partial}{\partial x^i}$,这里$\xi^i(x)$是$U$上的$C^\infty$函数. 这个切口称为$U$上的一个向量场. 同样,$T^*M$的切口是$\omega(x)=\lambda_i(x)\mathrm{d}x^i$,$\lambda_i(x)$是光滑的,它称为$U$上的一个微分形式. 注意,它们都是局部切口. 一个向量丛是否有整体切口存在,是一个问题. 又,有时我们把底

空间 $M$ 看成是 $E$ 的零切口.

**4. 浸入、嵌入、浸没和子流形** 微分流形的子流形当然是一个重要概念.其实,所有的微分流形都可以看成是某一个高维欧氏空间 $\mathbf{R}^N$ 的子流形,只不过因为这样做等于对每个微分流形都规定要放在一个包含空间中去研究,而包含空间的概念是不必要的.而且,一旦把微分流形 $M$ 放在 $\mathbf{R}^N$ 中去,$\mathbf{R}^N$ 自己就有一个标准的微分结构,这个微分结构又与 $M$ 的微分结构有什么关系呢?但不论如何,子流形的概念总是极其重要的,而且其关键问题就在于如上所述的两种微分结构的关系.所以我们应该从一个明确无误的定义开始.

**定义 6** 设 $M$ 是一 $n$ 维微分流形,$N\subset M$ 称为 $M$ 的一个 $k$ $(k\leqslant n)$ 维子流形,是指对每一点 $P\in N$ 都有一个在 $M$ 中的坐标邻域 $(U,\varphi)$ 使得

(1) $\varphi(N\cap U)=\mathbf{R}^k\cap\varphi(U)$;

(2) $(N\cap U,\varphi|_{N\cap U})$ 是 $P$ 在 $N$ 中的坐标邻域和坐标映射.

**注 1** 很容易证明,$\{N\cap U,\varphi|_{N\cap U}\}$ 是 $N$ 的一个图册,所以 $N$ 本身也成为一个微分流形,而其微分结构是由 $M$ 的微分结构诱导来的.

**注 2** 我们又时常定义一个映射——包含映射(inclusion) $i:N\rightarrow M$, $P(\in N)\mapsto P(\in M)$,于是子流形 $N$ 又可看成是包含映射的像:$N=i(N)$.于是子流形的定义又可以这样来看:首先我们已有了一个 $k$ 维微分流形 $N$,作为一个集合,它是 $M$ 的子集.因此可以认为 $i(N)=N$.如果 $i(N)$ 是如定义中说的 $M$ 的子流形,则 $M$ 在 $i(N)$ 上诱导出一个微分结构,所以,一方面 $N$ 与 $i(N)$ 是一一对应的,但是二者的关系又不止于此,$N$ 和 $i(N)$ 还有相同的微分结构:$N$ 本身自己已有一个微分结构,$i(N)$ 则从 $M$ 中继承了一个诱导的微分结构,二者应该相同.所以,$N$ 与 $i(N)$ 是微分同胚的(当然也是同胚的).这样把 $N$ 与 $i(N)$ 分开的好处在于下面我们将不必限制于 $i$ 而对某一可微映射 $f$ 把 $N$ 与 $f(N)$ 比照起来研究.

上述关于子流形的定义虽然在几何上是十分清楚自然的,用起来并不方便,因为它要求去寻找适当的坐标邻域和坐标映射 $(U,\varphi)$.解决这个困难的基本工具是反函数定理.

第三章§5 中我们讲了局部的反函数定理,即其中的定理 1,我们把它重新叙述于下:

设 $U\subset\mathbf{R}^n$ 是一开集,$f:U\rightarrow\mathbf{R}^n$ 是一个光滑映射,且 $0\in U$,$f(0)=0$,若 $(Tf)(0)$ 是非退化的,则 $f$ 必是 $x=0$ 在 $U$ 中的某个邻域到 $f(0)$ 在 $\mathbf{R}^n$ 中的某个邻域的微分同胚.

我们先把它用于微分流形的情况,于是把 $U$ 改成某个 $n$ 维微分流形 $M$,可微映射 $f:M\rightarrow N$, $P\in M$, $f(P)=Q\in N$,$N$ 是另一个同维数的微分流形,而 $T_Pf$ 是非退化的.这时很容易证明,$f$ 在 $P$ 附近是局部的微分同胚.这个证明是很容易的:只需取 $M$ 与 $N$ 在 $P$ 与 $Q$ 附近的坐标邻域 $(U,\varphi)$ 与 $(V,\psi)$,并用 $f$ 之坐标表示 $\psi\circ f\circ\varphi^{-1}$ 作为反函数定理中的 $f$,用 $\varphi(U)\subset\mathbf{R}^n$ 作为其中的 $U$,并以 $\psi(Q)=0$ 的邻域 $\mathbf{R}^n$ 作为其中的 $f(0)$ 之邻域即可.$(Tf)(0)$ 之非退化性现在就是 $\psi\circ f\circ\varphi^{-1}$ 在 0 处的雅可比行列式不为 0.

这个定理对于 $M$ 与 $N$ 之维数不相同:$\dim M=m$, $\dim N=n$ $(m\neq n)$ 的情况也有推广.这就是第三章§5 的定理 2 与定理 3.我们很容易地把它们用于微分流形的情况,并分别称之为浸入(immersion)定理与浸没(submersion)定理.下面我们直接对微分流形 $M$ 与 $N$ 的可微映射 $f$ 写出它们而不加证明.先给出两个定义,其中用到 $M,N,P,Q$ 等等都已如上述.

**定义 7** 当 $m\leqslant n$ 时,若 $T_Pf:T_PM\rightarrow T_QN$ 是一个单射,就说 $f$ 在 $P$ 点附近是一个浸入.若

$f$ 在 $M$ 之每一点附近均为浸入，就说 $f: M \to N$ 是一个浸入.

在通常的书上，常用条件 $m<n$，其实当 $m=n$ 时，由于 $T_Pf$ 是一个方阵，而一个方阵是单射与它是一个一一映射是一样的. 但有一点要说明：按第三章 §5 的定理 2 的说法，定义 7 应改为“$T_Pf$ 当 $P$ 在某点附近时为单射”，为什么这里只说在 $P$ 点为单射？用局部坐标表示最容易说明这一点. 现在 $T_Pf$ 是一个 $n \times m$ 矩阵，而单射条件，即此矩阵之秩为 $m$，于是有一个 $m$ 阶子行列式不为 0. 但是这个矩阵之元为光滑函数，所以其子行列式为连续函数，只要在 $P$ 点不为 0，则在 $P$ 点附近也不为 0. 又由于现在 $m$ 是最大秩，此矩阵之秩不会大于 $m$，所以一旦有一个 $m$ 阶子行列式在 $P$ 点不为 0，则它在 $P$ 附近之秩均为 $m$. 这就是在定义中只说 $T_Pf$ 在 $P$ 点为单射的理由. 下面讲满射时也有相类似的情况. 在那里我们就不再提醒这一点了.

**定理 10**（**局部浸入定理**） 设 $f: M \to N$ 在 $P \in M$ 处是一单射，这里 $m=\dim M<n=\dim N$，则必存在 $P$ 与 $f(P)=Q$ 的邻域中的局部坐标 $x$ 与 $y$，使 $f$ 可以表为

$$f(x^1, \cdots, x^m) = (x^1, \cdots, x^m; 0, \cdots, 0). \tag{32}$$

(32) 右方有时称为一个典则浸入 (canonical immersion)：它把 $\mathbf{R}^m$ 浸入在 $\mathbf{R}^n$ $(n>m)$ 时，而把 $\mathbf{R}^m$ 中的一点 $(x^1, \cdots, x^m)$ 变成 $\mathbf{R}^n$ 中 $(x^1, \cdots, x^m; 0, \cdots, 0)$，正如把 $\mathbf{R}^1$ 看成 $(x, y)$ 平面 $\mathbf{R}^2$ 中的 $x$ 轴那样.

浸入必然是 $M$ 与 $f(M)$ 的局部的微分同胚. 这使我们设想，$M$ 在 $N$ 中的浸入像 $f(M) \subset N$ 应该局部地是 $N$ 的子流形. 一个特别明显的例子是曲线. 若 $y=f(x)$ 是一条光滑曲线，它必可看作是由 $\mathbf{R}^1$ 到 $\mathbf{R}^2$ 的浸入像，这个浸入可以写为

$$x = t, y = f(t), \quad t \in (a, b).$$

$M$ 是 $\mathbf{R}$ 上的开区间 $(a, b)$，$N$ 则是 $\mathbf{R}^2$. 这个映射的切映射是

$$t \to (x'(t), f'(t)).$$

它自然是单射，因为 $x'(t) \equiv 1$. 稍加推广可以类似地讨论更高维的几何图形如曲面等等. 但是，浸入像并不一定整体是子流形. 首先，$f(M)$ 可能是自交的. 例如图 7-2-6 所示确实是一个浸入. 因为局部地看来，$(A_1, B_1)$ 被映入曲线上的上行弧段 $\overset{\frown}{AB}$，而把 $P_1$ 映到 $Q$，这是一个微分同胚，同样 $(C_1, D_1)$ 与下行弧段 $\overset{\frown}{CD}$ 也可以互相微分同胚. 但是 $f(\mathbf{R})$ 显然不是 $\mathbf{R}^2$ 的子流形. 即令 $f(M)$ 不自交而与 $M$ 可以是一对一的，它也不一定是子流形. $y=\sin \dfrac{1}{x}$，映 $M=(0, +\infty)$ 到 $\mathbf{R}^2$ 就是一个例子. 如果考虑曲面和其它更高维的情况就更复杂了. 但是因为至少局部地看来，浸入像是子流形，所以有时也称 $f(M)$ 为浸入子流形 (immersed submanifold). 为了使 $f(M)$ 成为如定义 6 那样的子流形，就需要对 $f$ 再加限制，例如要求 $f$ 是嵌入映射 (embedding)：

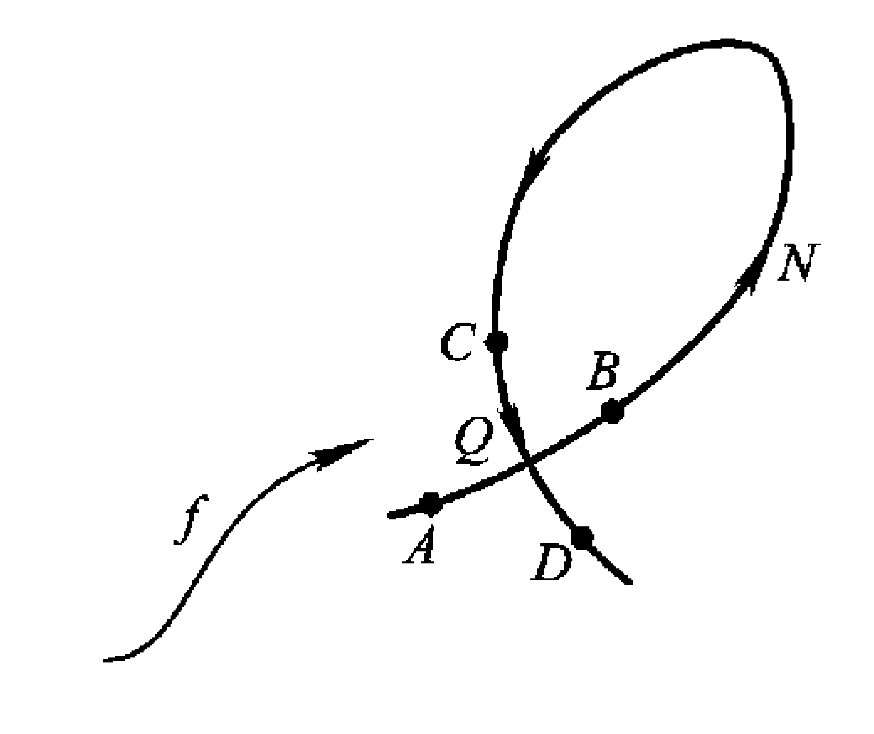


图 7-2-6

**定义 8** 若 $M, N$ 是两个微分流形而 $\dim M=m<\dim N=n$，$f: M \to N$ 是可微映射，如果

(1) $f$ 在 $M$ 之各点上均为浸入；

(2) $M$ 和 $f(M)$ 之间有一对一关系；

(3) $M$ 与赋有 $N$ 之诱导微分结构的 $f(M)$ 微分同胚；

则 $f$ 称为一个嵌入.

嵌入像当然是一个子流形.这类子流形称为嵌入子流形(embedded submanifold).

本节中一再提到一个抽象的 $k$ 维微分流形不一定能在 $\mathbf{R}^{k+1}$ 中实现,例如黎曼曲面就是这样.实际上,惠特尼(H. Whitney)证明了下面的定理.

**惠特尼嵌入和浸入定理** $k$ 维微分流形必可嵌入在 $\mathbf{R}^{2k+1}$ 中而成为其嵌入子流形;或者浸入在 $\mathbf{R}^{2k}$ 中而成为其浸入子流形.

读者会问,一个浸入在什么条件下成为嵌入?对这一类问题我们不说了.

上面我们讲了一种定义子流形的方法,即用一个低维流形向高维流形作映射,并且看映射的像是否子流形.但是还有另一个方法,即考虑一个高维流形向低维流形的映射,并且考虑低维流形的某一点在此映射下的原像,看它是否高维流形的子流形.前一个方法以浸入概念为基础,后一个方法则以浸没(submersion)概念为基础.所以我们首先给出浸没概念的定义.

**定义 9** 设 $f: M \to N$ 是两个微分流形之间的可微映射,$\dim M = m > \dim N = n$,$P \in M$,$f(P) = Q \in N$,若 $T_P M$ 为一满射就说 $f$ 在 $P$ 点为一个浸没(submersion),若 $f$ 在 $M$ 之每一点均为浸没就说 $f$ 在 $M$ 上为一浸没.

这里我们有

**定理 11(局部浸没定理)** 设上述 $f: M \to N$ 为一浸没,$f() = Q$,则必存在 $P$ 的邻域 $U$ 中的局部坐标 $(x^1, \cdots, x^m)$ 以及 $Q$ 在 $N$ 中的邻域 $V$ 中的局部坐标 $(y^1, \cdots, y^n)$,使在此坐标系下,映射 $f$ 可写为

$$y = f(x) = \pi(x^1, \cdots, x^m) = (x^1, \cdots, x^n), \tag{33}$$

这里 $\pi: \mathbf{R}^m \cong \mathbf{R}^n \times \mathbf{R}^{m-n} \to \mathbf{R}^n$ 是映 $\mathbf{R}^m$ 为坐标的前 $n$ 个分量 $(x^1, \cdots, x^n) \in \mathbf{R}^n$ 的投影映射.这个投影映射也称为一个典则浸没(canonical submersion).

这个定理就是第三章§5的定理3投影定理.利用它可以得到子流形的一个判据,但在这以前我们先引进一个概念.

**定义 10** 设 $f: M \to N$ 是微分流形 $M$ 到 $N$ 的可微映射.$\dim M = m$,$\dim N = n$.$P \in M$,$Q \in N$

(1) 如果 $T_P f$ 之秩小于 $n$,则称 $P$ 为 $f$ 之临界点(或称奇点),反之称 $P$ 为 $f$ 之正则点.

(2) 若 $P$ 为 $f$ 之临界点,则称 $f(P)$ 为 $f$ 的临界值,若 $f^{-1}(Q)$ 中一切点均为正则点,则称 $Q$ 为 $f$ 之正则值.

这样,若对一切 $P \in f^{-1}(Q)$,$T_P f$ 是浸没,则 $Q$ 是正则值.有了这样一个概念,我们要把隐函数定理(第三章§5定理4)稍微改动一下:

对 $f, M, N$ 仍作如上的解释.

**隐函数存在定理** 令 $m > n$,若 $Q \in f(M)$ 是 $f$ 的正则值,则 $f^{-1}(Q)$ 是一个微分流形,其微分结构是由 $M$ 的微分结构诱导而来,而且 $\dim f^{-1}(Q) = m - n$,或者说 $n$ 是 $f^{-1}(Q)$ 的余维数.

其实证明是很简单的,我们不妨设 $P, Q$ 在某局部坐标中均为原点.第三章的定理是说 $f^{-1}(Q)$ 中有某一点 $P$(也设它为原点),使 $T_P f$ 之秩为 $n$(但设源空间的维数为 $m+n$,所以自然有 $m+n>n$),这时 $T_P f$ 自然是满射,而在 $P$ 附近,按局部浸没定理,$f$ 可以写为

$$f_j(x^1, \cdots, x^m; x^{m+1}, \cdots, x^{m+n}) = x^j, j = 1, 2, \cdots, m.$$

如果我们再取一个新的局部坐标

$$z^j = f_j(x) - x^j, \qquad j = 1,2,\cdots,m,$$
$$z^{m+j} = x^{m+j}, \qquad j = 1,2,\cdots,n,$$

则 $f^{-1}(0)$在 $P$ 附近就成为 $\mathbf{R}^n$：

$$\mathbf{R}^n: z^j = 0, \qquad j = 1,2,\cdots,m.$$

现在 $M$ 中的微分结构是 $\mathbf{R}^{n+m}$，而 $f^{-1}(0)$在 $P$ 附近的微分结构是 $\mathbf{R}^n$，它自然是由 $\mathbf{R}^{n+m}$诱导而来.所以是子流形结构.所以，第三章中的隐函数定理适用于 $f^{-1}(Q)$之一点附近.用现在的提法，使用正则值的概念，这个证明适用于 $f^{-1}(Q)$中的一切点.又因 $f^{-1}(Q)\subset M$，令 $i$ 为包含映射，$i[f^{-1}(Q)]=f^{-1}(Q)$，所以子流形的一切要求 $f^{-1}(Q)$均可适合.这样就有关于子流形的一个最常用的判据：

**定理 12(原像定理)**　若 $Q$ 为 $f: M\to N$ 之正则值，则 $f^{-1}(Q)$为一个维数为 $m-n$ 的子流形.

这个定理与我们所熟悉的称子流形为以下方程组之解

$$f^j(x^1,\cdots,x^m) = 0, j = 1,2,\cdots,n < m$$

相一致.这里要求这些方程是“独立的”，即雅可比矩阵$\dfrac{\partial(f^1,\cdots,f^n)}{\partial(x^1,\cdots,x^m)}$在这个方程组的轨迹上有最大秩 $n$.

现在把这个定理用于上面一个没有解决的问题：$O(n)$是不是一个微分流形？设 $A$ 为一 $n$ 阶方阵，所谓 $A\in O(n)$即 $A$ 应适合${}^tA\cdot A=I$，${}^tA$ 是$A$ 的转置矩阵.但不论 $A$ 是否正交矩阵，${}^tA\cdot A$总是对称方阵，记一切 $n$ 阶对称方阵之集为 $S(n)$，我们已说过，一个 $n$ 阶方阵可以看成 $\mathbf{R}^{n^2}$ 中的一个点，其各元素即为此点之坐标.$S(n)$于是也可看成 $\mathbf{R}^{\frac{1}{2}n(n+1)}$.于是 $O(n)$就是以下方程在 $\mathbf{R}^{n^2}$ 中的解集合：

$$f(A) \equiv {}^tA\cdot A = I. \tag{34}$$

我们视 $f$ 为 $\mathbf{R}^{n^2}$ 到 $S(n)\cong\mathbf{R}^{\frac{1}{2}n(n+1)}$ 的映射，或者

$$O(n) = f^{-1}(I).$$

现在以 $\mathbf{R}^{n^2}$ 为 $M$，$S(n)$为 $N$，我们需要证明的就只是 $I$ 为$f$ 的正则值.注意到 $\mathbf{R}^{n^2}$ 与 $S(n)$都是 $\mathbf{R}^k$ 形的微分流形，所以 $T_P\mathbf{R}^{n^2}\cong\mathbf{R}^{n^2}$，$T_QS(n)=S(n)$.所以只要证明 $f$ 之切映射在$f^{-1}(I)$为满射即可.但由切映射之局部坐标形式，并且注意到一个矩阵的元素即是其各个坐标分量，我们通过计算 $f(x+h)-f(x)$来求出 $\mathrm{d}f$，这里 $x=A$，不过 $h$ 应与$x$ 同维数，所以我们令 $x=A$，$h=B$ 都是$n$ 阶方阵，计算 $\mathrm{d}f$ 时应略去阶数为 1 以上的量，现在则略去含 $B$ 之元素平方的量，这样得到

$$\begin{aligned} f(A+B)-f(A) &= ({}^tA+{}^tB)(A+B)-{}^tA\cdot A \\ &= {}^tA\cdot B+{}^tBA+{}^tBB, \end{aligned}$$

略去最后一项即得

$$\mathrm{d}_Af\cdot B = {}^tA\cdot B+{}^tB\cdot A. \tag{35}$$

所以为了证明 $\mathrm{d}_Af:\mathbf{R}^{n^2}(\cong T_A\mathbf{R}^{n^2})\to S(n)(\simeq T_{f(A)}S(n))$为满射，只需证明对任意 $D\in S(n)$，

一定可以找到 $B\in\mathbf{R}^{n^2}$ 使

$$^tA\cdot B+{}^tB\cdot A=D.\tag{36}$$

因为 $D\in S(n)$，$D={}^tD$，所以上式右方可写为$\dfrac{D}{2}+\dfrac{^tD}{2}$而我们来解

$$^tA\cdot B=\frac{1}{2}D.$$

因为 $A$ 适合$^tA\cdot A=I$ 所以$^tA$ 非奇异，故由上式

$$B=\frac{1}{2}(^tA)^{-1}D.\tag{37}$$

但这个 $B$ 必适合

$$^tB=\frac{1}{2}{}^tD\cdot A^{-1}.$$

故

$$^tB\cdot A=\frac{1}{2}{}^tD=\frac{D}{2}.$$

而(37)给出的 $B$ 即适合(36)，所以 $\mathrm{d}f_A$ 在 $f^{-1}(I)$上是满射. 应用原像定理即知 $O(n)$是一个 $n^2-\dfrac{1}{2}n(n+1)=\dfrac{1}{2}n(n-1)$维微分流形.

这个微分流形由两个连通分支组成. 其中一个是 $SO(n)$，另一个之元行列式为 $-1$.

# §3　多重线性代数介绍

向量丛的结构是非常丰富的，以微分流形每一点切空间与余切空间为基础，可以建立许多非常重要的数学对象，其中用得最多的是张量和多重向量. 我们需要从这些对象的代数结构上来观察它们. 这就是多重线性代数. 本节中我们介绍它的两个最重要的部分——张量代数与外代数.

**1. 张量代数**　§1 中我们已经通过一些实例介绍了什么是张量. 在那里我们是在空间的某个区域中讨论一般的坐标变换，例如 $y=y(x)$. 但是在微分学中我们已指出了这种一般的非线性变换局部地可以看成是线性空间中的线性变换. §2 中就指出，可以在一点 $P$ 的切空间 $T_PM$ 上用切变换来代替它，所以应该先在线性空间框架中彻底地研究它们. 所以又引入了 $T\mathbf{R}^n$ 上的线性变换. 其实我们已经初步地这样做了，我们的处理方式是通过张量的分量来了解张量：在每一个坐标系下指定一组 $n^r$ 个实数，并要求在不同坐标系下的相应的 $n^r$ 实数组有适当的关系. 即是说不同坐标系下的分量适合相当的变换关系. 这样的讲法当然有很大的好处，例如便于计算. 但即使对最简单的张量——即向量，这种讲法与通常微积分教材中对向量的处理就有一些区别. 我们不妨再一次概括一下这种讲法. 向量，亦即一阶张量我们分成两大类，第一类就是适合线性空间的运算法则的对象称为向量. 这些运算法则的几何含意十分清楚，即平行四边形法则. 第二类向量就是前一类向量的线性函数，或在用第六章的语言来说就是其线性泛函. 但是线性泛函也是线性运算的对象，所以也是向量. 不过它所组成的线性空间与第一类向量组成的空间 $V\simeq T\mathbf{R}^n$ 不同，我们称它为 $V$ 之对偶空间，记作 $V^*\simeq(T\mathbf{R}^n)^*$. 这一类向量却缺少那么丰富的直观

背景，而且因为线性代数理论告诉我们，$(T\mathbf{R}^n)^*$ 与 $T\mathbf{R}^n$ 同构，那么二者之间的区别何在呢？为了弄清这个问题就发现当把向量用分量来表示时就很明白了．从笛卡儿建立解析几何起我们就明白了，坐标系是研究几何对象的好方法．所以在研究向量时，我们首先在 $T\mathbf{R}^n$ 中取一个基底，也叫做取一个标架．所谓基底就是 $n$ 个线性无关的向量 $e_i\in T\mathbf{R}^n$，我们记为 $\{e_1,\cdots,e_n\}$，于是有了任一向量 $u\in T\mathbf{R}^n$，一定有 $n$ 个实数 $u^i\in T\mathbf{R}$ 使 $u=u^ie_i$，这 $n$ 个实数就叫做 $u$ 对上述基底的坐标或分量，我们用 $(u^1,\cdots,u^n)$ 表示．现在我们作一个规定，凡是基底一定用花括号 $\{\cdots\}$ 表示，所以 $\{\cdots\}$ 内的对象都是向量，凡是坐标一定用圆括号 $(\cdots)$ 表示，其中的对象一定是数．我们在 §1 中讲向量时是用的与坐标无关的处理方法．所以要讨论在标架变换下，坐标如何变换．如果旧标架是 $\{e_1,\cdots,e_n\}$，新标架是 $\{f_1,\cdots,f_n\}$，而有

$$f_i = a_i^j e_j. \tag{1}$$

或者写成

$$\{f\} = A\{e\}.$$

$A=(a_i^j)$，而 (1) 是对上标求和．同一个向量 $u$ 如在不同的标架下有不同的坐标 $(u^1,\cdots,u^n)$ 与 $(v^1,\cdots,v^n)$，则应有

$$u = v^i f_i = u^i e_i$$

将 (1) 代入上式即得

$$u^j = a_i^j v^i \tag{2}$$

或

$$(v) = ({}^tA)^{-1}(u).$$

在 §1 中我们说 (1) 与 (2) 是互相逆步的，所以标架的变换与坐标的变换是互相逆步的．这是一个基本的事实．

现在转到对偶空间，如果原空间 $V$ 的标架为 $\{e\}$，后来变成了 $\{f\}$，其对偶空间 $V^*$ 之元，如果后来是 $\{\sigma\}$ 而原来是 $\{\omega\}$，则应有

$$\langle\omega,e\rangle = \langle\sigma,f\rangle$$

这里 $e,f;\omega,\sigma$ 都是向量而不是实数．以 (1) 代入上式，并用转置矩阵处理有

$$\langle\omega,A^{-1}f\rangle = \langle {}^tA^{-1}\omega,f\rangle = \langle\sigma,f\rangle$$

因此

$$\sigma = {}^tA^{-1}\omega. \tag{3}$$

它与 (1) 互相逆步，而与 (2) 相一致，$V^*$ 中的元应该用 $\{\omega\}=\{\omega^1,\cdots,\omega^n\}$ 或 $\{\sigma\}=\{\sigma^1,\cdots,\sigma^n\}$ 来表示，如果同一个 $u^*\in V^*$ 表示为

$$u^* = \lambda_i\omega^i = \mu_i\sigma^i,$$

又容易得到

$$\mu_i = a_i^j\lambda_j \text{ 即 } (\mu) = A(\lambda). \tag{4}$$

它与 (1) 一致而与 (2) 逆步．§1 中我们给出了逆变向量与协变向量的概念．其实是逆变还是协变要看对谁而言．如果以 $V$ 中标架变换公式 (1) 为准，凡与它一致的就称为协变，与它逆步的就称为逆变，则以上所述可以列成一个表

| | 原空间 $V$ | 对偶空间 $V^*$ |
|---|---|---|
| 标架 | (1) 基准 | (3) 与(1)逆步 |
| 坐标 | (2) 逆变坐标 | (4) 与(1)一致,协变坐标 |

我们还有一个非常方便的记号:凡用上标表示的均与(1)逆步地变换,凡用下标表示的均与(1)一致地变换.

我们现在希望对张量也能这样地作与坐标无关的处理.这样做的好处是可以对线性代数与解析几何的一些根本问题有更进一步的理解.而且因为张量不再如向量那么直观,作抽象的处理反而更容易.

§1 中我们举了两类张量的例子.一类是例如向量积,角动量,另一类如黎曼度量.严格地说第一类还不是一般的张量,而是所谓 $p$ 向量,亦即外积,正是我们下面要讲的.所以现在我们从黎曼度量

$$ds^2 = g_{ij}dx^i dx^j \tag{5}$$

开始.§1 中我们已经说过了 $(g_{ij})$ 是一个二阶协变张量,$dx^i$ 则是一个一阶逆变张量,但是最好我们把(5)式右方的 $dx^i$,$dx^j$ 分开来看,认为它们是独立的,类似的情况我们在第三章中也就看过.那里我们在讲高阶微分时曾问过,如果在讲一阶微分 $df = Ah$ 时,我们用 $h$ 表示切向量,再求第二次微分时应看到 $A$ 其实依赖于 $x$,所以 $d(df) = dA\cdot h = Bh\cdot h$.$B$ 在局部坐标下是 $f(x)$ 的二阶偏导数所成的矩阵 $\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right)_0$ 即黑塞矩阵(Hessian matrix).下标 0 表示在某一点 $x_0$ 计算它.可是为什么计算 $dA = Bh$ 中要用相同的 $h$ 呢?其实应该用 $k$.$k$ 并不是本质上与 $h$ 不同,二者同为任意的切向量,只不过要强调 $k$ 与 $h$ 的独立性而已.所以二阶微分可以看成是

$$\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right)_0 hk = (B; h, k) \tag{6}$$

即 $B$ 作用于两个互相独立的切向量 $h$ 和 $k$.我们这里采用了与向量的内积相似的记号,因为它们实在是相似的:一阶微分是 $A$ 与 $h$ 的内积 $df = (A, h)$,对 $h$ 是线性的,二阶微分是 $B$ 与 $h$ 和 $k$ 的"内积"$(B; h, k)$,它对 $h$ 与 $k$ 分别是线性的,所以称为双线性形式.不同之处则在于 $A$ 是协变向量,$B$ 却不是二阶协变张量——这当然有深刻的原因,即曲率问题.现在回到黎曼度量.在一些很老的历史文献中,都是把 $dx^i$ 与 $dx^j$ 分开来的,前者记为 $dx$,后者记为 $\delta x$,而有

$$ds^2 = g_{ij}dx^i \delta x^j.$$

现代的文献中当然已不再这样写,但是老记号给了我们一个启发:怎样来认识张量 $(g_{ij})$?它是由 $dx^i$ 和 $\delta x^j$(它们都在空间 $V$ 中,现在 $V = T_pM$,所以它们是切向量而 $dx^i$,$\delta x^i$ 是一些数,是切向量的逆变坐标.)到 $\mathbf{R}$ 的映射.因为 $dx^i$ 与 $\delta x^j$ 是互为独立的,所以 $ds^2$ 是由乘积空间 $V\times V$ $(= T_pM \times T_pM)$ 到 $\mathbf{R}$ 的映射.这个映射是多重线性(multilinear)的.所谓一个映射

$$T: V\times V \to \mathbf{R}, \quad (u, v) \mapsto T(u, v)$$

为多重线性的,即指,当 $u$ 或者 $v$ 固定时,我们得到对另一个变元的线性映射,即

$$\begin{aligned} T(\lambda_1 u_1 + \lambda_2 u_2, v) &= \lambda_1 T(u_1, v) + \lambda_2 T(u_2, v), \\ T(u, \mu_1 v_1 + \mu_2 v_2) &= \mu_1 T(u, v_1) + \mu_2 T(u, v_2). \end{aligned} \tag{7}$$

这样我们就给出

**定义 1**　一个由 $V\times V$ 到 $\mathbf{R}$ 的双线性(bilinear)映射:

$$T: V\times V\to\mathbf{R}\tag{8}$$

称为二阶协变张量(或(0,2)型张量).所有二阶协变张量形成一个线性空间,记作 $\mathscr{T}_2^0(V)$.

定义中指出 $\mathscr{T}_2^0(V)$ 是一个线性空间,其实这是有待证明的,只不过因为它是如此明显,所以我们未加证明而在定义中直接指出这一点.于是要问,$\dim\mathscr{T}_2^0(V)=$? 我们可以证明 $\dim\mathscr{T}_2^0(V)=n^2$,而且具体地找出 $n^2$ 个线性无关的二阶协变张量.为此目的,我们还要再引入两个向量的张量积.设 $\omega,\sigma\in V^*$,我们可以定义其张量积为一个二阶协变张量如下:任取 $u,v\in V$,我们定义

$$(\omega\otimes\sigma)(u,v)=\omega(u)\cdot\sigma(v).\tag{9}$$

很明显它是 $V\times V$ 到 $\mathbf{R}$ 的双线性映射:

**定义 2**　双线性映射(9)称为 $\omega$ 与 $\sigma$ 之张量积,记作 $\omega\otimes\sigma$.

现在我们就很容易找出 $\mathscr{T}_2^0(V)$ 的一个基底了.令 $\{\omega^1,\cdots,\omega^n\}$ 为 $V^*$ 的一个基底,我们要证明 $\{\omega^i\otimes\omega^j\}$,$i,j=1,\cdots,n$,就符合要求,为此证明它们是线性无关的.设 $C_{ij}\omega^i\otimes\omega^j=0$,令 $\{e_1,\cdots,e_n\}$ 为 $\{\omega^1,\cdots,\omega^n\}$ 之对偶基底,取 $u=e_{i_0}$,$v=e_{j_0}$,以 $C_{ij}\omega^i\otimes\omega^j$ 作用于 $(u,v)\in V\times V$,有

$$0=(C_{ij}\omega^i\otimes\omega^j)(e_{i_0},e_{j_0})=C_{i_0j_0}.$$

所以一切 $C_{ij}=0$,而知 $\{\omega^i\otimes\omega^j\}$ 在 $\mathscr{T}_2^0(V)$ 中线无关.其次易见所有二张协变张量 $T$ 均可写成

$$T=T_{ij}\omega^i\otimes\omega^j.\tag{10}$$

实际上,只要取 $T_{ij}=T(e_i,e_j)$ 即可.由此可知 $\{\omega^i\otimes\omega^j\}$ 确为 $\mathscr{T}_2^0(V)$ 的一个基底,而且 $\dim\mathscr{T}_2^0(V)=n^2$.

现在我们要问,如果在 $V$ 中作一个坐标变换,即由老基底 $\{e_1,\cdots,e_n\}$ 变为新基底 $\{f_1,\cdots,f_n\}$,变换为

$$f_i=a_i^je_j$$

问 $T$ 应如何改变?一方面看来 $T$ 应该不变,因为按定义,$T: V\times V\to\mathbf{R}$ 是一个双线性变换,这是与坐标是无关的.但是 $\{\omega^i\}$ 会改变,$(T_{ij})$ 也会改变,实际上,$\{\omega^i\}$ 作为 $V^*$ 之基底(即标架),若在新坐标系下成为 $\{\Omega^j\}$,则应有

$$\omega^i=a_j^i\Omega^j.$$

它与 $\{e_1,\cdots,e_n\}$ 的变换是互相逆步的.若 $T$ 在新坐标下为 $(\tilde{T}_{kl})$,则由(10)式

$$T=T_{kl}\omega^k\otimes\omega^l=T_{kl}a_i^ka_j^l\Omega^i\otimes\Omega^j=\tilde{T}_{ij}\Omega^i\otimes\Omega^j.$$

所以

$$\tilde{T}_{ij}=T_{kl}a_i^ka_j^l.\tag{11}$$

所以 $T_{ij}$ 的变化规律和协变向量的变化规则相符.在§1中,我们对张量是用分量来定义的:在某一个坐标系下,一个二阶协变张量即一组 $n^2$ 个实数 $(T_{ij})$,如果在另一个坐标系下,这 $n^2$ 个实数成为 $(\tilde{T}_{kl})$,则二者之间应有关系式(11).这样,我们就看到,现在给出的 $\mathscr{T}_2^0(V)$ 张量都是§1中讲的协变张量.反过来,如果给出一个§1中讲的二阶协变张量 $(T_{ij})$,令

$$T = T_{ij}\omega^i \otimes \omega^j$$

则得到我们现在讲的 $\mathscr{T}_2^0(V)$ 张量，而且§1 中给出的分量变化规则(6)，保证了(10)式给出的 $T$ 确实是一个与坐标无关的由 $V\times V$ 到 $\mathbf{R}$ 的双线性映射，因此也是现在讲的 $\mathscr{T}_2^0(V)$ 张量. 总之，§1 的用分量讲的张量与现在作为双线性映射的张量是完全一样的. 现在的讲法能帮助我们更好地了解张量的本质，而§1 的讲法计算起来更加便利.

以上我们考虑了最常见的二阶协变张量，它的定义显然是 $V^*$ 之元即协变向量之推广. 所以协变向量就是一阶协变张量. 其余类型的张量自可仿此处理：我们仍考虑一个 $n$ 维线性空间 $V$，但同时也考虑其对偶空间 $V^*$，其维数仍为 $n$，如果有 $r$ 个 $V^*$，$s$ 个 $V$，我们要研究由 $\underbrace{V^*\times\cdots\times V^*}_{r\text{个}}\times\underbrace{V\times\cdots\times V}_{s\text{个}}$ 到 $\mathbf{R}$ 的多重线性映射，于是有

**定义 3** 上述多重线性映射

$$\underbrace{V^*\times\cdots\times V^*}_{r\text{个}}\times\underbrace{V\times\cdots\times V}_{s\text{个}}\to\mathbf{R} \tag{12}$$

称为一个 $(r,s)$ 型张量（$r$ 阶逆变，$s$ 阶协变），这种张量构成一个线性空间，记作 $\mathscr{T}_s^r(V)$.

和前面一样，若 $V$ 和 $V^*$ 有对偶基底 $\{\omega^i\}$，$\{\sigma_j\}$，$i,j=1,2,\cdots,n$，则对 $\underbrace{V^*\times\cdots\times V^*}_{r\text{ 个}}\times\underbrace{V\times\cdots\times V}_{s\text{ 个}}$ 中的元 $\{\mu_1,\cdots,\mu_r;\nu^1,\cdots,\nu^s\}$ 我们可以定义张量积

$$\begin{aligned}&\omega^{i_1}\otimes\cdots\otimes\omega^{i_r}\otimes\sigma_{j_1}\otimes\cdots\otimes\sigma_{j_s}\{\mu_1,\cdots,\mu_r;\nu^1,\cdots,\nu^s\}\\ =\ &\omega^{i_1}(\mu_1)\cdots\omega^{i_r}(\mu_r)\sigma_{j_1}(\nu^1)\cdots\sigma_{j_s}(\nu^s),\end{aligned} \tag{13}$$

而知一切 $\mathscr{T}_s^r(V)$ 张量 $t$ 均可写为

$$t = t_{i_1\cdots i_r}^{j_1\cdots j_s}\omega^{i_1}\otimes\cdots\otimes\omega^{i_r}\otimes\sigma_{j_1}\otimes\cdots\otimes\sigma_{j_s}. \tag{14}$$

于是 $\{\prod_{l=1}^{r}\omega^{i_l}\prod_{l=1}^{s}\sigma_{j_l}\}$ 是 $\mathscr{T}_s^r(\overline{V})$ 的基底，从而

$$\dim\mathscr{T}_s^r(V) = n^{r+s}. \tag{15}$$

$t_{j_1\cdots j_s}^{i_1\cdots i_r}$ 是 $t$ 的分量. 我们时常除了(14)式以外，也用下面的记号：

$$t = t_{j_1\cdots j_s}^{i_1\cdots i_r}. \tag{16}$$

它表示现在讲的张量与§1 中讲的张量完全一样，因此，§1 中讲的张量运算的法则——特别是缩并——在此全部适用，我们就不再重复了.

但是现在在张量中增加了一种新的代数运算，即张量积. 在最简单的情况下张量积由(9)决定，一般由(13)给出. 因此很容易看出，张量积是不可交换的. 例如由(9)就可以看到，

$$(\sigma\otimes\omega)(\mu,\nu) = \sigma(\mu)\cdot\omega(\nu),$$

而与 $(\omega\otimes\sigma)(\mu,\nu)$ 不同. 但是，张量积很容易看到是结合的，而且适合分配律.

现在把关于张量的概念用到微分流形上去. 设 $M$ 是一个微分流形，于是对每一点 $P\in M$，有一个切空间 $T_PM$ 以及余切空间 $T_P^*M$，二者互为对偶空间. 如果 $M$ 上有一个坐标邻域 $U$. $P\in U\subset M$ 有局部坐标 $(x^1,\cdots,x^n)$. 则 $T_PM$ 和 $T_P^*M$ 分别有基底 $\left\{\frac{\partial}{\partial x^1},\cdots,\frac{\partial}{\partial x^n}\right\}$ 和 $\{\mathrm{d}x^1,\cdots,$

$\mathrm{d}x^n\}$. 我们现在以 $T_PM$ 为 $V\cong\mathbf{R}^n$，则可以定义其上的 $(r,s)$ 型张量

$$t_P = t^{i_1\cdots i_r}_{j_1\cdots j_s}(P)\frac{\partial}{\partial x^{i_1}}\otimes\cdots\otimes\frac{\partial}{\partial x^{i_r}}\otimes\mathrm{d}x^{j_1}\otimes\cdots\otimes\mathrm{d}x^{j_s}\Big|_P. \tag{17}$$

$\{t_P\}$构成一个线性空间 $\mathscr{T}^r_s(P)$. 于是令 $P$ 变动时，将可得到一个丛，即 $M$ 上的张量丛. 它有一个全空间 $E=\bigcup_{P\in M}\mathscr{T}^r_s(P)$. 有一个底空间即 $M$，有由 $E$ 到 $M$ 的投影 $\pi: E\to M, \mathscr{T}^r_s(P)$，也有纤维 $\pi^{-1}(P)=\mathscr{T}^r_s(P)$. 这样一个张量丛不但有一般的微分流形结构，而且有向量丛的结构，特别是局部平凡性，即存在一个同构：

$$\bigcup_{P\in M}t_p\simeq U\times\mathscr{T}^r_s\simeq U\times\mathbf{R}^{n^{r-s}}.$$

这一点和切丛、余切丛是一样的，即是说，所谓局部平凡性即在 $M$ 的一个适当的坐标邻域中各点 $P$ 处纤维 $\pi^{-1}(P)$ 具有相同的基底. 但是纤维作为线性空间的“系数”是 $x$ 的函数. 所以在研究这类问题时 $M$ 上 $U$ 中的 $C^\infty$ 函数环——以后记作 $\mathscr{F}$——将有特别作用. 对这个张量丛我们也可以定义其 $C^\infty$“切口”($C^\infty$ section)，简单说即在每一个 $\pi^{-1}(P)$ 中取一固定的元，其局部表示是(17)，不过 $t^{i_1\cdots i_r}_{j_1\cdots j_s}(P)$ 是 $C^\infty$ 函数，即为 $\mathscr{F}$ 中的元. 上面我们只是说明了建立张量丛的程序与建立切丛、余切丛一样. 但是要实现这个程序，特别是在讨论其局部平凡化时，还有不少计算，在此我们全部略去. 不过有一点必须提到：在考虑某一定点 $P$ 处的纤维 $\pi^{-1}(P)$ 时，我们得到的 $\mathscr{T}^r_s(V)$ 确实是线性空间，因为(17)的系数均为常数；但在讨论 $C^\infty$ 切口及这类切口的集合时，(17)的“系数”都成了 $\mathscr{F}$ 中的元. 注意，线性空间的系数是一个“域”中的元，本书中除了少数特殊情况都是以实数域 $\mathbf{R}$ 作为“系数”域(应该称为基域)，而 $\mathscr{F}$ 却只是一个环. 这样得到的代数结构与线性空间是很不相同的，称为一个模(module). 在研究 $M$ 上的张量丛时就会遇到这方面的问题. 我们还要提醒的是：局部平凡化的结果使纤维与底空间的坐标完全分离，所以在(17)中，底空间的坐标只出现在系数 $t^{i_1\cdots i_r}_{j_1\cdots j_s}(P)$ 中，而$\frac{\partial}{\partial x^{i_1}}\otimes\cdots\otimes\mathrm{d}x^{j_s}$ 完全与之无关. 这样，如果要对局部坐标求导数，只需在 $t^{i_1\cdots i_r}_{j_1\cdots j_s}(P)$ 中施行即可.

最后我们再讲一点关于记号的问题. 我们是由黎曼的度量公式(5)引入度量张量$(g_{ij})$的，但是按上面讲的张量理论，(5)式应如何理解却成了问题. 其实，在较古老的文献例如黎曼本人的著作中，$\mathrm{d}s^2$ 了解为无穷接近的两点之间的距离. 涉及无穷小自然会带来许多麻烦，因此在当代的文献中都是说，在 $M$ 之每点的切空间 $T_PM$ 上引入一个内积，而要求它光滑地依赖于 $P$. 有了内积，即可对任一切向量 $X_P\in T_PM$ 赋以一个非负的数 $(X,X)$，即其长度的平方. $X$ 既然是线性空间中的向量，自然无所谓无穷小，而所谓光滑地依赖于 $P$ 是指光滑地依赖于底空间的坐标. 所以情况与讨论微分学时把 $x+h$ 中的 $x$ 与 $h$ 严格区别开来作类似. 所谓内积可由一个正定双线性型来刻画，第六章 §2 定义 3 中已详细地讲到. 因此，我们在 $T_PM$ 上有了一个双线性映射

$$T_PM\times T_PM\to\mathbf{R},\qquad X,Y\in T_PM\mapsto(X,Y)\in\mathbf{R}.$$

按现在的讲法就是有一个二阶协变张量 $g_{ij}$ 如果在 $T_PM$ 中引入标准的基底$\left\{\frac{\partial}{\partial x^1},\cdots,\frac{\partial}{\partial x^n}\right\}$，从而

在 $T_P^*M$ 中有了相应的对偶基 $\{\mathrm{d}x^1,\cdots,\mathrm{d}x^n\}$，则此张量为

$$T = g_{ij}\mathrm{d}x^i \otimes \mathrm{d}x^j,\qquad (X,X) = (T;X,X) \tag{18}$$

现在的 $\{\mathrm{d}x^1,\cdots,\mathrm{d}x^n\}$ 是 $T_P^*M$ 的基底，所以 $\mathrm{d}x^i$ 是向量，而与(5)式中的 $\mathrm{d}x^i$ 作为切向量 $(\mathrm{d}x^1,\cdots,\mathrm{d}x^n)$ 的分量，是一些数，二者不同. 以(18)作用到 $\left(\frac{\partial}{\partial x^k},\frac{\partial}{\partial x^l}\right)\in T_PM\times T_PM$ 上去即得

$$\left(T:\frac{\partial}{\partial x^k},\frac{\partial}{\partial x^l}\right) = \left(\frac{\partial}{\partial x^k},\frac{\partial}{\partial x^l}\right) = g_{kl}.$$

说这个内积光滑地依赖于 $P$ 即指 $g_{ij}$ 是 $x$ 的 $C^\infty$ 函数. 如果我们取一个"无穷小"切向量 $(\mathrm{d}x^1,\cdots,\mathrm{d}x^n) = \mathrm{d}x^i\cdot\frac{\partial}{\partial x^i}$（$\mathrm{d}x^i$ 是分量是数，$\frac{\partial}{\partial x^i}$ 是基底向量是切向量），立即得到(5)式. 于是我们得到了两个式子(18)与(5). (18)是张量的表达式，其中 $\mathrm{d}x^i$ 是一个余切向量，是 $T_P^*M$ 中的基底 $\{\mathrm{d}x^1,\cdots,\mathrm{d}x^n\}$ 中的第 $i$ 个元素. (5)则是这个张量作用到切向量 $X=\mathrm{d}x^i\cdot\frac{\partial}{\partial x^i}$ 的结果，这里 $\mathrm{d}x^i$ 是 $X$ 的第 $i$ 个分量，它是一个数，而不是向量. 如果把它与作为余切向量的 $\mathrm{d}x^i$ 混起来，会发生很大的混淆. 总之，黎曼原来使用的记号(5)与现在我们使用的(18)，尽管很相似，却是不相同的. 现在的数学家们心里都很明白这个区别，但是 $\mathrm{d}x^i$ 这些记号用熟了，也就不太在乎，所以有的文献干脆把(1)写成

$$\mathrm{d}s^2 = g_{ij}(x)\mathrm{d}x^i \otimes \mathrm{d}x^j.$$

但是，事关"无穷小量"就更应该说明. 其实，一个合适的讲法是，在切向量之长度定义为 $(X,X)^{\frac{1}{2}}$ 以后，如果要求图 7－3－1中曲线弧 $\widehat{AB}$ 之长为 $s$，这里 $\widehat{AB}$ 的方程是 $x^i = x^i(t)$ 即 $\boldsymbol{x}=\boldsymbol{x}(t)$，$A$ 点对应于参数 $t=0$，$B$ 点对应于参数 $t$，则把 $\widehat{AB}$ 分成若干段，而例如 $\widehat{A_{i-1}A_i}$ 则代以弦 $\overline{A_{i-1}A_i}$，然后再代以 $A_{i-1}$ 处（即 $t=t_{i-1}$ 处）的切线，因此弦将被切向量 $\dot{\boldsymbol{x}}(t_{i-1})(t_i-t_{i-1})=\dot{\boldsymbol{x}}(t_{i-1})\Delta_i t$ 取代，其长为 $\sqrt{\langle\dot{\boldsymbol{x}}(t_{i-1}),\dot{\boldsymbol{x}}(t_{i-1})\rangle}\cdot(t_i-t_{i-1})$ $(t_i\geqslant t_{i-1})$. 把分点 $\{A_i\}$ 取密以求极限可得

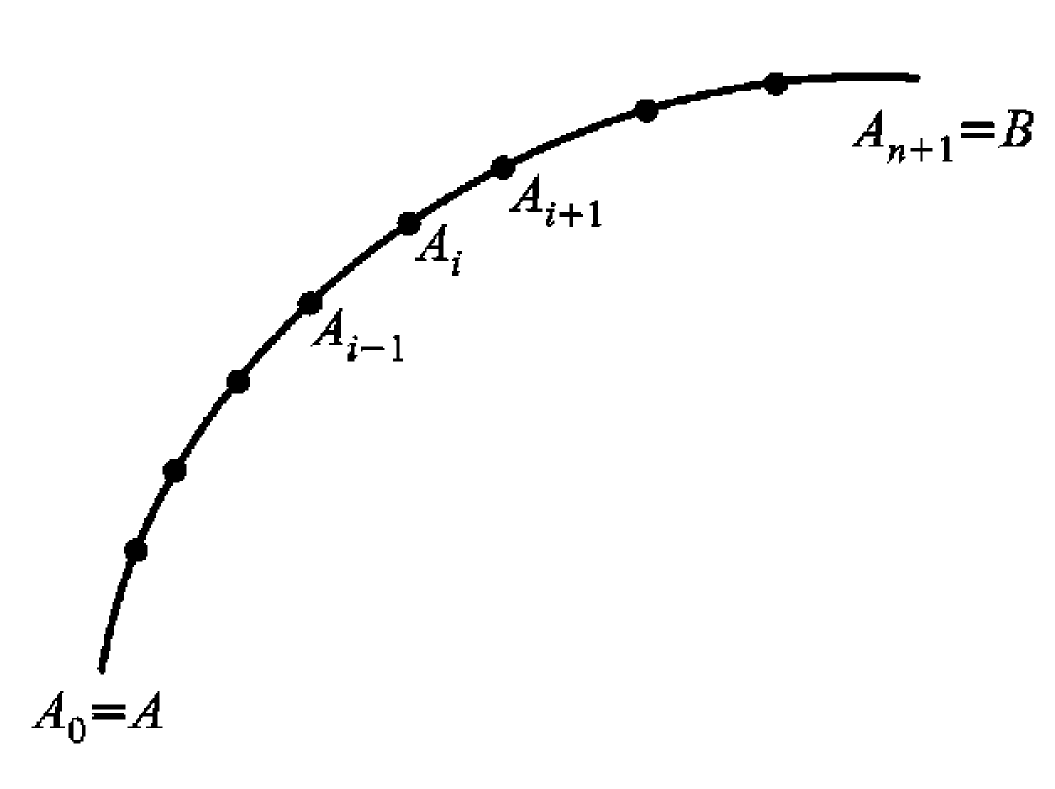



图 7－3－1

$$s = \int_0^t \sqrt{\langle\dot{\boldsymbol{x}}(t),\dot{\boldsymbol{x}}(t)\rangle}\,\mathrm{d}t.$$

所以

$$\mathrm{d}s^2 = \langle\dot{\boldsymbol{x}}(t),\dot{\boldsymbol{x}}(t)\rangle\mathrm{d}t^2 = g_{ij}(x)\mathrm{d}x^i\mathrm{d}x^j.$$

这就是通过在 $M$ 上给出切向量长度 $(X,X)$ 以及作典型的积分运算得到黎曼度量公式(1)的方法. 这里没有因为无穷小量的介入而导致的任何困难.

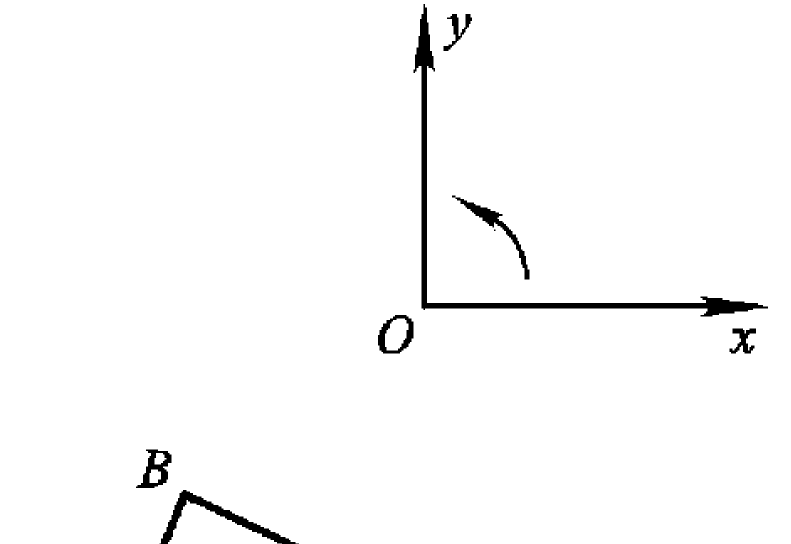
在一本著名的微分几何数本（M. Spivak，Comprehensive Introduction to Differential Geometry，Vol. 1，Publish or Perish，1979），153 页上说到这类记号问题时讲了一段很有趣的话："古典的微分几何学家（还有古典分析学家）在讲到坐标 $x^i$ 的无穷小改变 $\mathrm{d}x^i$ 时，毫不犹豫. 莱布尼茨就是这样. 他们谁也不肯承认这种说法是毫无意义的，因为把这些无穷小量相除（只要做得对）总

会给出正确结果.最终人们认识到,最接近实情的描述无穷小变化的办法就是描述这个变化的方向,亦即切向量.原来,$\mathrm{d}f$ 被看作是在点的无穷小变化下函数 $f$ 的变化,现在,$\mathrm{d}f$ 必须看成是这个变化的函数,亦即切向量的函数(就是说 $\mathrm{d}f$ 要通过 $\mathrm{d}f(X)=X(f)$ 才能体现出来,所以 $\mathrm{d}f$ 作为 $X$ 的"函数"就是说它是 $T_PM$ 的对偶空间 $T_P^*M$ 之元——本书作者注),然后 $\mathrm{d}x^i$ 本身也摇身一变成为切向量$\frac{\partial}{\partial x^i}$的函数(即 $\mathrm{d}x^i\left(\frac{\partial}{\partial x^j}\right)=\delta_j^i$ 是 $\mathrm{d}x^i$ 之定义——本书作者注),认识到这一点以后,剩下来要做的只有一件事:搞上一些新定义,但是仍然使用老记号,再让读者慢慢来跟.总之,所有涉及无穷小的古典的概念,都像 $\mathrm{d}f$ 一样,变成了切向量的函数,唯一例外是无穷小量的商,因为它们自己就变成了切向量$\frac{\mathrm{d}c}{\mathrm{d}t}$($c$ 就是这里讲的曲线 $\boldsymbol{x}=\boldsymbol{x}(t)$——本书作者注)".

本章所讲的几乎都是老概念、老记号的新理解,新定义.我们所做的一切都是为了帮助"读者慢慢来跟".

**2. 外代数**　19 世纪中叶,人们开始研究高维空间,这不但是由于数学本身的需要,而且许多力学、物理学问题迫切需要它.于是线性代数理论出现了.矩阵、向量、超复数、四元数……可以说使人眼花缭乱.其实都是围绕一个问题:高维空间的几何量要怎样来刻画才好.这个问题的重要性是不言而喻的.当时一位著名数学家格拉斯曼(H. Grassmann,1809—1877)作了伟大的贡献.他指出,各种各样几何量原来是互有联系的系统,他的表示方式也是系统的.例如,以$(x_1,y_1)$,$(x_2,y_2)$ $(x_3,y_3)$为顶点的三角形面积是

$$\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$



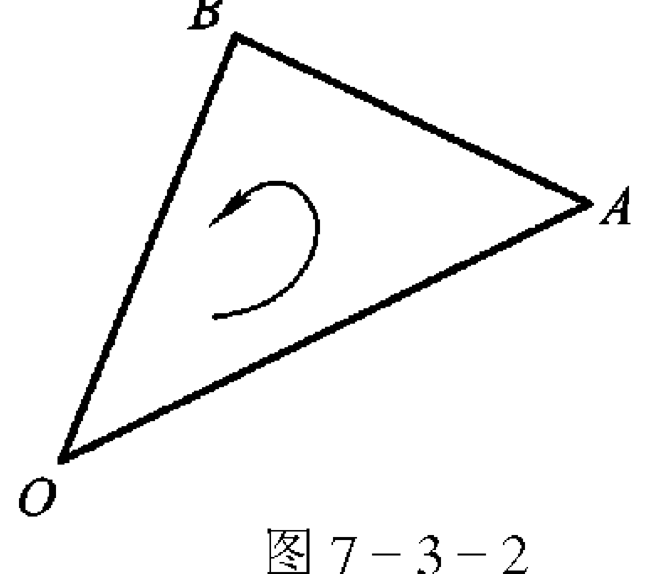




图 7-3-2

如果不管因子$\frac{1}{2}$,并且只看其矩阵,则其二阶子矩阵$\begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{pmatrix}$的三个二阶子行列式

$$X=x_1-x_2,\qquad Y=y_1-y_2,\qquad Z=x_1y_2-x_2y_1$$

分别是有向线段$\overrightarrow{(x_1,y_1),(x_2,y_2)}$的两个投影以及原点 $O$ 与这两点所成三角形的有向面积的 2 倍.既然面积与长度是从同一个根源中来,自然也应该有符号.

线段长有符号,人们都容易接受.面积有符号,在大多数微积分教科书中却不承认.面积的符号与坐标系采取何种定向有关.例如设 $x$ 轴与 $y$ 轴成为右手系,则由$\overrightarrow{OA}$到$\overrightarrow{OB}$形成的三角形$\triangle OAB$,如果$\overrightarrow{OA}$与$\overrightarrow{OB}$也成右手系,则其面积算是正的;由$\overrightarrow{OB}$到$\overrightarrow{OA}$形成的三角形面积就算是负的(图 7-3-2).这一个规定会为整个数学的发展带来很大的好处.坚持这一点的数学家中就有默比乌斯,而默比乌斯带的发现也与此密切相关(1865).格拉斯曼另一个十分深刻的思想是:应该与坐标的选取无关地探讨几何量的性质,这正是本章的根本指导思想.

从常用的微积分教本中看到,上述三角形之面积是一个向量$\overrightarrow{OA}\times\overrightarrow{OB}$,我们自然会问,它是逆变向量还是协变向量?于是取两个标架且设 $A$,$B$ 两点在这些标架即坐标系中为$(a^1,a^2)$,$(b^1,b^2)$以及$(\tilde{a}^1,\tilde{a}^2)$,$(\tilde{b}^1,\tilde{b}^2)$.如果标架变换用(1)表示,则经过简单计算即知在新旧坐标系下

之面积适合

$$\frac{1}{2}\begin{vmatrix}\tilde{a}^1 & \tilde{a}^1\\ \tilde{b}^2 & \tilde{b}^2\end{vmatrix} = \det A \cdot \frac{1}{2}\begin{vmatrix}a^1 & a^2\\ b^1 & b^2\end{vmatrix}$$

可见面积这个几何量与向量本质不同.

其实,说面积是一个向量,即向量积$\overrightarrow{OA}\times\overrightarrow{OB}$还有一个限制,即这个说法最多只能用于 $\mathbf{R}^3$. 因为若$\overrightarrow{OA}$,$\overrightarrow{OB}$是 $n$ 维向量:$\overrightarrow{OA}=(a^1,\cdots,a^n)$,$\overrightarrow{OB}=(b^1,\cdots,b^n)$,则当 $n=3$ 时,$\overrightarrow{OA}\times\overrightarrow{OB}$的分量恰好是下述 $2\times3$ 矩阵

$$\begin{pmatrix}a^1 & a^2 & \cdots & a^n\\ b^1 & b^2 & \cdots & b^n\end{pmatrix}$$

的三个二阶子行列式(加上适当的 ± 号). 但是对于一般的 $n$,上述矩阵共有$\frac{1}{2}n(n-1)$个二阶子行列式,为使它们能够组成一个 $n$ 维向量,必须$\frac{1}{2}n(n-1)=n$,即 $n=3$. 可见非 3 维的 $\mathbf{R}^n$ 中是不可能定义两个向量的向量积的. 通常的微积分教材在讲向量分析与场论时总限于在 $\mathbf{R}^3$ 上(有的书上未强调这一点)原因在此. 但这些二阶子行列式有几何意义正是格拉斯曼的基本思想的一部分.

那么,用什么样的数学工具来表示面积、体积呢? 用格拉斯曼建立的外代数. 下面我们就来介绍这个理论. 我们要提醒读者经常把下面的结果与 $n$ 维平行体的体积和行列式作比较. 其实如上所述行列式从几何上看正是"表示"了 $n$ 维平行体的体积. 它是一个十分重要的概念,而且又是外代数最好的模型.

我们再回到张量空间. 为了简单起见,我们恒取 $V=\mathbf{R}^n$ 以后主要是讨论协变张量,但照顾到外代数习惯用上标记号 $\wedge^r$,我们也就把 $\mathscr{T}_r^0$ 型张量暂时记作 $\mathscr{T}^r$. 下面我们讲张量恒指它研究过的张量有两类特别重要的,一是对称张量,一是反对称(skew symmetric)张量. 设有张量 $T\in\mathscr{T}^r(\mathbf{R}^r)$,于是对 $r$ 个 $n$ 维向量 $v_1,\cdots,v_r$ 有

$$T(v_1,\cdots,v_r)\in\mathbf{R}.$$

如果将 $v_i$ 与 $v_j$ 对换,时而会有

$$T(v_1,\cdots,v_j,\cdots,v_i,\cdots,v_r) = T(v_1,\cdots,v_i,\cdots,v_j,\cdots,v_k),\tag{19a}$$

时而又会有

$$T(v_1,\cdots,v_j,\cdots,v_i,\cdots,v_k) = -T(v_1,\cdots,v_i,\cdots,v_j,\cdots,v_k).\tag{19b}$$

相应地就说 $T$ 对于指标 $i$ 和 $j$ 是对称或反对称的. 如果 $T$ 对一切 $(i,j)$ 均是对称或反对称的,就说它是对称张量与反对称张量. 例如黎曼度量张量就是一个二阶对称协变张量,而反对称张量正是我们要讨论的主题.

如果把一个 $n$ 阶行列式 $\det A$ 的各行均看成一个 $n$ 维向量,则行列式显然是一个 $n$ 阶张量 $T$ 作用于这 $n$ 个 $n$ 维向量 $v_i=(a_{i1},\cdots,a_{in})$,$i=1,2,\cdots,n$(其实按我们的规定应该写成 $v_i=(a_i^1,\cdots,a_i^n)$,现在我们是照顾熟知的行列式记号)之值

$$\det A = T(v_i,\cdots,v_n),$$

它就是一个反对称张量作用于这 $n$ 个向量之值(互换行列式之两行将使之变号). 用初等方式定

义行列式时要用到置换算子 $\sigma$，每个置换(substitution，permatation)都是若干个对换(transposition)之积，如果是奇数(偶数)个，则称为奇(偶)置换，从而可以定义 $\sigma$ 之符号函数

$$\operatorname{sgn}\sigma=\begin{cases}+1, & \text{若 } \sigma \text{ 为偶置换;}\\ -1, & \text{若 } \sigma \text{ 为奇置换.}\end{cases}$$

有了 $\sigma$ 和 $\operatorname{sgn}\sigma$ 以后，张量的对称性和反对称性可以陈述如下：

**定义 4**　设 $T\in\mathscr{T}^r(\mathbf{R}^n)$，如果对任意置换 $\sigma$ 以及任意 $r$ 个向量 $v_1,\cdots,v_r$，分别有

$$T(v_1,\cdots,v_r)=T(v_{\sigma(1)},\cdots,v_{\sigma(r)}),\tag{20a}$$

$$T(v_1,\cdots,v_r)=\operatorname{sgn}\sigma\cdot T(v_{\sigma(1)},\cdots,v_{\sigma(r)}),\tag{20b}$$

就相应称 $T$ 为对称与反对称张量.

(19)和(20)的一致性是明显的.对称张量我们暂置一旁，重要的是反对称张量.所有 $r$ 阶反对称张量构成 $r$ 阶张量作为线性空间的子空间记为 $\wedge^r(\mathbf{R}^n)$：$\wedge^r(\mathbf{R}^n)\subset\mathscr{T}^r(\mathbf{R}^n)$.每一个 $r$ 阶张量均可反对称化，具体说我们要引进一个反对称化(或称交代化)算子(alternating operator)

$$Alt:\mathscr{T}^r(\mathbf{R}^n)\to\wedge^r(\mathbf{R}^n),$$

$$Alt(T)(v_1,\cdots,v_r)=\frac{1}{r!}\sum_{\sigma}\operatorname{sgn}\sigma T(v_{\sigma(1)},\cdots,v_{\sigma(r)}),\tag{21}$$

而有

**定理 1**　$Alt:\mathscr{T}^r(V)\to\wedge^r(V)$是一个投影算子.

**证**　我们要证明的是：$Alt$ 是一个满射，而且 $Alt^2=Alt$.先要注意，如果 $T$ 已经是反对称算子了，则 $Alt(T)=T$.事实上这时由(20)式知(21)之每一项均等于 $T(v_1,\cdots,v_r)$，因此

$$Alt(T)(v_1,\cdots,v_r)=\frac{1}{r!}\sum_{\sigma}T(v_1,\cdots,v_r)=T(v_1,\cdots,v_r).\tag{22}$$

这就是说，在 $\wedge^r(\mathbf{R}^n)$上，$Alt$ 其实是恒等算子：$Alt^2=Alt$.(21)式中要有一个因子 $\frac{1}{r!}$ 的作用就在于此.

再证 $Alt$ 是满射，为此先证对任意 $T\in\mathscr{T}^r(\mathbf{R}^n)$，$Alt(T)\in\wedge^r(\mathbf{R}^n)$.为此取任意置换 $\tau$ 以及 $r$ 个向量 $v_1,\cdots,v_r$.从简单的代数考虑知道 $\operatorname{sgn}(\sigma\tau)=\operatorname{sgn}\sigma\cdot\operatorname{sgn}\tau$，所以

$$\begin{aligned}Alt(T)(v_{\tau(1)},\cdots,v_{\tau(r)})&=\frac{1}{r!}\sum_{\sigma}\operatorname{sgn}\sigma T(v_{\sigma\tau(1)},\cdots,v_{\sigma\tau(r)})\\&=\frac{\operatorname{sgn}\tau}{r!}\sum_{\sigma'=\sigma\tau}\operatorname{sgn}\sigma' T(v_{\sigma'(1)},\cdots,v_{\sigma'(r)})=\operatorname{sgn}\tau\cdot AltT(v_1,\cdots,v_r).\end{aligned}$$

所以 $Alt(T)\in\wedge^r(\mathbf{R}^n)$，或写作 $Alt(\mathscr{T}^r(\mathbf{R}^n))\subset\wedge^r(\mathbf{R}^n)$.但是(22)告诉我们，在 $\wedge^r(\mathbf{R}^n)$上，$Alt$ 实际上是恒等算子，所以

$$\wedge^r(\mathbf{R}^n)=Alt\ \wedge^r(\mathbf{R}^n)\subset Alt(\mathscr{T}^r(\mathbf{R}^n))\subset\wedge^r(\mathbf{R}^n).$$

而此式中所有的“$\subset$”，均可写为“$=$”.特别是

$$Alt\mathscr{T}^r(\mathbf{R}^n))=\wedge^r(\mathbf{R}).$$

这样就得到 $Alt$ 是满射而定理证毕.

$\wedge^r(\mathbf{R}^n)$ 作为线性空间是 $\mathscr{T}^r(\mathbf{R}^n)$ 的子空间，这是几乎自明的事.下面略去其证明.但是张量还有一种乘法运算：张量积 $\otimes$，若 $T\in\wedge^r(\mathbf{R}^n)$，$S\in\wedge^s(\mathbf{R}^n)$，$T\otimes S$ 显然不一定仍是反对称的.

这就是说，$\bigotimes\limits_{r}\wedge^{r}(\mathbf{R}^{n})$ 不是 $\bigotimes\limits_{r}\mathscr{T}^{r}(\mathbf{R}^{n})$ 的子代数.我们还需要另一种乘法：外积.其定义如下：

**定义 5**　若 $T\in\wedge^{r}(\mathbf{R}^{n})$，$S\in\wedge^{s}(\mathbf{R}^{n})$，定义其外积 $T\wedge S\in\wedge^{r+s}(\mathbf{R}^{n})$ 如下：

$$T\wedge S=\frac{(r+s)!}{r!s!}Alt(T\otimes S).\tag{23}$$

**定理 2**　外积是双线性的、结合的.

**证**　外积的双线性很容易证，在此略去.我们来证明其结合性.除了定义 5 中的 $T,S$ 外，再加上 $R\in\wedge^{t}(\mathbf{R}^{n})$，我们来证明

$$T\wedge(S\wedge R)=(T\wedge S)\wedge R.$$

为此我们把记号简化一下：用 $\sigma$ 表示 $(1,2,\cdots,r+s+t)$ 的置换，用 $\tau$ 表示 $(1,2,\cdots,r+s+t)$ 中最后 $t$ 个数不动的置换，即只在 $(1,2,\cdots,r+s)$ 中作置换，指标为 $r+s+j,j=1,2,\cdots,t$，的都不变.这样 $\tau$ 与 $(1,2,\cdots,r+s)$ 之置换 $\sigma'$ 具有相同的符号函数：$\mathrm{sgn}\tau=\mathrm{sgn}\sigma'$.于是

$$\begin{aligned}
&Alt[Alt(T\otimes S)\otimes R](v_1,\cdots,v_{r+s+t})\\
&=\frac{1}{(r+s+t)!}\sum_{\sigma}\mathrm{sgn}\,\sigma\,Alt(T\otimes S)(v_{\sigma(1)},\cdots,v_{\sigma(s+t)})\\
&\quad\cdot R(v_{\sigma(r+s+1)},\cdots,v_{\sigma(r+s+t)})\\
&=\frac{1}{(r+s+t)!}\,\frac{1}{(r+s)!}\sum_{\sigma,\tau}\mathrm{sgn}\,\sigma\mathrm{sgn}\,\tau\cdot T(v_{\sigma\tau(1)},\cdots,v_{\sigma\tau(r)})\\
&\quad\cdot S(v_{\sigma\tau(r+1)},\cdots,v_{\sigma\tau(r+s)})\cdot R(v_{\sigma(r+s+1)},\cdots,v_{\sigma(r+s+t)})\\
&=\frac{1}{(r+s+t)!}\,\frac{1}{(r+s)!}\sum_{\tau}\sum_{\sigma}\mathrm{sgn}\,\varphi T(v_{\varphi(1)},\cdots,v_{\varphi(r)})\\
&\quad\cdot S(v_{\varphi(r+1)},\cdots,v_{\varphi(r+s)})\cdot R(v_{\varphi(r+s+1)},\cdots,v_{\varphi(r+s+t)}).
\end{aligned}$$

这里我们记 $\varphi=\sigma\tau$，而因 $\tau$ 保持 $(r+s+1,\cdots,r+s+t)$ 之次序不变，故 $\sigma(r+s+j)=\sigma\tau(r+s+j)$，$j=1,\cdots,t$.但是上式中 $\sigma$ 遍取 $(1,2,\cdots,r+s+t)$ 之一切排列时，$\varphi$ 也遍取 $(1,2,\cdots,r+s+t)$ 之一切排列，所以上式进一步又可化为

$$\begin{aligned}
&\frac{1}{(r+s)!}\sum_{\tau}\frac{1}{(r+s+t)!}\sum_{\varphi}\mathrm{sgn}\,\varphi\,T(v_{\varphi(1)},\cdots,v_{\varphi(r)})\\
&\quad\cdot S(v_{\varphi(r+1)},\cdots,v_{\varphi(r+s)})\cdot R(v_{\varphi(r+s+1)},\cdots,v_{\varphi(r+s+t)})\\
&=\frac{1}{(r+s)!}\sum_{\tau}Alt(T\otimes S\otimes R)(v_1,\cdots,v_{r+s+t})\\
&=Alt(T\otimes S\otimes R)(v_1,\cdots,v_{r+s+t}).
\end{aligned}$$

所以

$$\begin{aligned}
Alt(T\otimes S\otimes R)&=Alt[Alt(T\otimes S)\otimes R]\\
&=\frac{r!s!}{(r+s)!}Alt[(T\wedge S)\otimes R]\\
&=\frac{(r+s)!t!}{(r+s+t)!}\cdot\frac{r!s!}{(r+s)!}(T\wedge S)\wedge R]\\
&=\frac{r!s!t!}{(r+s+t)!}(T\wedge S)\wedge R.
\end{aligned}$$

同样的推理得知,它也应该等于$\frac{r!s!t!}{(r+s+t)!}T\wedge(S\wedge R)$,所以

$$(T\wedge S)\wedge R=T\wedge(S\wedge R).$$

我们可以把这个公共值定义为$T\wedge S\wedge R$,由此还可以进一步得到

**推论3** 设$T_i\in\wedge^{r_i}(\mathbf{R}^n)$,$i=1,2,\cdots,k$则可以定义$T_1\wedge\cdots\wedge T_k\in\wedge^{r_1+\cdots+r_k}(\mathbf{R}^n)$,而且

$$T_1\wedge T_2\wedge\cdots\wedge T_k=\frac{(r_1+r_2+\cdots+r_k)!}{r_1!r_2!\cdots r_k!}Alt[T_1\otimes\cdots\otimes T_k]. \tag{24}$$

定理2证明了外积是结合的,那么它是不是可交换的?实际上它既不如$\boldsymbol{A}\times\boldsymbol{B}$那样是反交换的:$\boldsymbol{B}\times\boldsymbol{A}=-\boldsymbol{A}\times\boldsymbol{B}$,更不是交换的,而有

**定理4** 若$T\in\wedge^r(\mathbf{R}^n)$,$S\in\wedge^s(\mathbf{R}^n)$,则

$$T\wedge S=(-1)^{rs}S\wedge T. \tag{25}$$

**证** 我们只需证明

$$Alt(T\otimes S)=(-1)^{rs}Alt(S\otimes T)$$

即可.为此取$r+s$个向量$v_1,\cdots,v_r,v_{r+1},\cdots,v_{r+s}$,有

$$Alt(T\otimes S)(v_1,\cdots,v_{r+s})=\frac{1}{(r+s)!}\sum_\sigma \operatorname{sgn}\sigma T(v_{\sigma(1)},\cdots,v_{\sigma(r)})\cdot S(v_{\sigma(r+1)},\cdots,V_{\sigma(r+s)}).$$

考虑一个特殊的置换$\tau$:$(1,\cdots,r,r+1,\cdots,r+s)\to(r+1,\cdots,r+s,1,\cdots,r)$,即将后$s$个元换成前$r$个元.例如把$r+1$变成1,为此$r+1$需要跨过$1,\cdots,r$等$r$个元,因而需要$r$个对换,$r+2$变成2,这又要$r$个对换,直到$r+s$变成$s$,共需$rs$个对换,故$\operatorname{sgn}\tau=(-1)^{rs}$,而上式成为

$$\begin{aligned}Alt(T\otimes S)(v_1,\cdots,v_{r+s})&=\frac{1}{(r+s)!}\sum_\sigma \operatorname{sgn}\sigma(\operatorname{sgn}\tau)^2S(v_{\sigma(r+1)},\cdots,v_{\sigma(r+s)})\cdot T(v_{\sigma(1)},\cdots,v_{\sigma(r)})\\&=\frac{\operatorname{sgn}\tau}{(r+s)!}\sum_{\sigma\tau}\operatorname{sgn}(\sigma\tau)S(v_{\sigma\tau(1)},\cdots,v_{\sigma\tau(s)})\cdot T(v_{\sigma\tau(s+1)},\cdots,v_{\sigma\tau(r+s)})\\&=(-1)^{rs}\frac{1}{(r+s)!}\sum_{\sigma'}\operatorname{sgn}\sigma'S(v_{\sigma'(1)},\cdots,v_{\sigma'(s)})T(v_{\sigma'(s+1)},\cdots,v_{\sigma'(r+s)})\\&=(-1)^{rs}Alt(S\otimes T)(v_1,\cdots,v_{r+s}).\end{aligned}$$

定理证毕.

外积之分配律是自明的:

$$T\wedge(S_1+S_2)=T\wedge S_1+T\wedge S_2,\ (T_1+T_2)\wedge S=T_1\wedge S+T_2\wedge S. \tag{26}$$

至此我们看到,$\wedge^r(\mathbf{R}^n)$是一个线性空间,而$\bigcup\limits_r\wedge^r(\mathbf{R}^n)$对于外积成为一个代数.它服从结合律,但不服从交换律(但又不是反交换的),现在余下的是求出$\wedge^r(\mathbf{R}^n)$的维数和基底.这里也附带提到,$\wedge^0(\mathbf{R}^n)=\mathscr{T}^0(\mathbf{R}^n)\simeq\mathbf{R}$.关于$\wedge^r(\mathbf{R}^n)$的维数和基底我们有

**定理5** 当$r>n$时,$\wedge^r(\mathbf{R}^n)=\{0\}$,即为零维空间,若$0\leqslant r\leqslant n$,则$\dim\wedge^r(\mathbf{R}^n)=\binom{n}{r}$,而若$\{\omega^1,\cdots,\omega^n\}$为$\wedge^1(\mathbf{R}^n)=\mathscr{T}^1(\mathbf{R}^n)$之基底,则$\Lambda^r(\mathbf{R}^n)$的基底是

$$\{\omega^{i_1}\wedge\cdots\wedge\omega^{i_r},1\leqslant i_1<i_2<\cdots<i_r\leqslant n\}. \tag{27}$$

证 令$\{e_1,\cdots,e_n\}$为$\mathbf{R}^n$的一个基底. 若$r>n$,任取$e_{i_1},\cdots,e_{e_{i_r}}$,其中必有重复者. 例如$e_{i_1}=e_{i_2}$,这时考虑$Alt\ T(e_{i_1},\cdots,e_{i_r})$,因为$Alt\ T$是反对称的,故$Alt\ T(e_{i_1},e_{i_2},e_{i_3},\cdots,e_{i_r})=-Alt\ T(e_{i_2},e_{i_1},e_{i_3},\cdots,e_{i_r})$,而当$e_{i_1}=e_{i_2}$时知它为0. 所以$\wedge^r(\mathbf{R}^n)$之一切元素均为0,而$\dim\wedge^r(\mathbf{R}^n)=0$.

今设$0\leqslant r\leqslant n$,$\mathcal{T}^r(\mathbf{R}^n)$之基底是$\{\omega^{i_1}\otimes\cdots\otimes\omega^{i_r}\}$这里$\{\omega^1,\cdots,\omega^n\}\subset(\mathbf{R}^n)^*$是$\{e_1,\cdots,e_n\}$的对偶基. 于是$\wedge^r(\mathbf{R}^n)$之元必为$Alt(\omega^{i_1}\otimes\cdots\otimes\omega^{i_r})$之线性组合,现在把$\{i_1,\cdots,i_r\}$按大小重排为$j_1<j_2<\cdots<j_r$(若$i_1,\cdots,i_r$中有相同者前已得知$Alt(\omega^{i_1}\otimes\cdots\otimes\omega^{i_r})=0$). 而$Alt(\omega^{i_1}\otimes\cdots\otimes\omega^{i_r})=\pm Alt(\omega^{j_1}\otimes\cdots\otimes\omega^{i_r})$. 这样对任意$T\in\mathcal{T}^r(\mathbf{R}^n)$,$Alt\ T$必可表示为$\{\omega^{j_1}\wedge\cdots\wedge\omega^{j_r}\}$,$j_1<j_2<\cdots<j_r$之线性组合. 即是说$\wedge^r(\mathbf{R}^n)$由$\{\omega^{j_1}\wedge\cdots\wedge\omega^{j_r}\}$,$j_1<j_2<\cdots<j_r$张成. 但是$\{\omega^{j_1}\wedge\cdots\wedge\omega^{j_r}\}$是线性无关的. 因为若有$a_{j_1\cdots j_r}\omega^{j_1}\wedge\cdots\wedge\omega^{j_r}=0$,把这个张量作用到$\{e_{k_1},\cdots,e_{k_r}\}$上去,即有$a_{k_1\cdots k_r}=0$. 所以$\{\omega^{j_1}\wedge\cdots\wedge\omega^{j_r}\}$是基底. 这个基底中元素的个数是由$n$个事物中取$r$个的组合数$\binom{n}{r}$($\mathcal{T}^r(\mathbf{R}^n)$之一个基底是$\{\omega^{i_1}\otimes\cdots\otimes\omega^{i_r}\}$,这里$i_1,\cdots,i_r$次序不受限制. 所以其中所含元素的个数是$n$中取$r$的排列数$(n)_r=n(n-1)\cdots(n-r+1)=r!\binom{n}{r}$). 至此$\dim\wedge^r(\mathbf{R}^n)=\binom{n}{r}$. 其实,证明的前一部分,也可归入这里,因为当$r>n$时,由定义$\binom{n}{r}=0$.

再回来看代数$\bigcup_r\wedge^r(\mathbf{R}^n)$. 现在知道它应写为$\wedge(\mathbf{R}^n)=\bigoplus_{r=0}^{n}\wedge^r(\mathbf{R}^n)$,而其维数为$\sum_{r=0}^{n}\binom{n}{r}=2^n$. 这个代数称为$\mathbf{R}^n$上的外代数或格拉斯曼代数(Grassmann algebra),$\wedge(\mathbf{R}^n)$中的$\wedge^r(\mathbf{R}^n)$之元称为$r$向量.

这里附带还可以解决一个问题. 在$\mathbf{R}^n$中任取$r$个向量,问它们在什么时候线性相关?因为$\mathbf{R}^n\cong(\mathbf{R}^n)^*$,不妨认为这个$r$个向量$\omega^1,\cdots,\omega^r\in(\mathbf{R}^n)^*$即是某个$n$维空间的对偶空间之元. 于是可以作出$\omega^1\wedge\cdots\wedge\omega^r$. 如果它们是线性相关的,则例如$\omega^r=\sum_{k=1}^{r-1}C_k\omega^k$,则

$$\omega^1\wedge\cdots\wedge\omega^r=\sum_{k=1}^{r-1}C_k\omega^1\wedge\cdots\wedge\omega^{r-1}\wedge\omega^k.$$

右方每一项中$\omega^k$至少出现两次,但由上面已证明$\omega^k\wedge\omega^k=0$,知上式右方每一项均为0,所以$\omega^1\wedge\cdots\wedge\omega^r=0$. 反之,设它们是线性无关的,则由线性代数中的定理,一定可以把它们扩充成为$(\mathbf{R}^n)^*$的一个基底. 于是$\omega^1\wedge\cdots\wedge\omega^r$是$\wedge^r(\mathbf{R}^n)$基底的一个元,它当然不会为0,由此得到

**推论6** $n$维向量$\omega^1,\cdots,\omega^r$为线性相关的充分必要条件是

$$\omega^1\wedge\cdots\wedge\omega^r=0. \tag{28}$$

其实这里的证明与行列式为0的充要条件是其各行线性相关的证明是一致的;也与$n$个向量线性相关的充分必要条件是其构成的$n$维平行体的体积为0是一致的.

至此要问,我们在格拉斯曼所指出的道路上走了多远?按拉格斯曼的思想,我们已将各种几

何量构成了格拉斯曼代数 $\wedge(\mathbf{R}^n)=\bigoplus_{r=0}^{n}\wedge^r(\mathbf{R}^n)$. 它是一个分级代数：$\wedge^0(\mathbf{R}^n)$ 就是实数域 $\mathbf{R}$，$\wedge^1(\mathbf{R}^n)$ 就是向量空间 $\mathbf{R}^n$（我们暂时不区分协变与逆变向量，因为二者之区别无非其一在对偶空间中，而 $(\mathbf{R}^n)^*\cong\mathbf{R}^n$，$\wedge^1(\mathbf{R}^n)$ 中的元可以用 $1\times n$ 矩阵表示：

$$G_n^1=(x^1,x^2,\cdots,x^n)$$

它的每一个分量即 $1\times1$ 子式恰好是这个向量在各个坐标轴上的投影. 再看 $\wedge^2(\mathbf{R}^n)$，它的元是以下形式的 2 向量之线性组合：$u^1\wedge u^2$，如果 $u^i$ 的分量是 $(u_i^1,\cdots,u_i^n)$，则这个 2 向量可以用 $2\times n$ 矩阵

$$G_n^2=\begin{pmatrix}u_1^1 & u_1^2 & \cdots & u_1^n\\ u_2^1 & u_2^2 & \cdots & u_2^n\end{pmatrix}$$

来表示，它的所有的 2 阶子式 $\begin{vmatrix}u_1^i & u_1^j\\ u_2^i & u_2^j\end{vmatrix}=u_1^iu_2^j-u_2^iu_1^j$ 恰好是 $u^1,u^2$ 这两个向量是在 $x^1x^2$ 坐标平面上的投影形成的平行四边形的面积，而我们可以把 $G_n^2$ 与下式对应

$$G_n^2\cong\sum_{i<j}(u_1^iu_2^j-u_2^iu_1^j)e_i\wedge e_j.$$

所以我们就说 $G_n^2$ 是 $u^1,u^2$ 这两个向量所成的平行四边形之面积. 读者不要以为这只是一个方便的说法. $u^1,u^2$ 尽管是 $n$ 维向量，但是以它们为基底（当然我们假设 $u^1,u^2$ 不平行，否则上面的讨论就会落了空），确实成为一个 2 维空间. 它确实是平面而不只是像平面. 我们中学里学过的平面几何定理（例如三角形的重心定理）在此全都成立，只不过有一些定理需要引入欧氏结构才能证明，如勾股定理. $\wedge_n^2(\mathbf{R}^n)$ 的表示法与通常微积分教本中讲的向量 $\overrightarrow{OA}=(a_1,a_2,a_3)$ 与 $\overrightarrow{OB}=(b_1,b_2,b_3)$ 之向量积 $\overrightarrow{OA}\times\overrightarrow{OB}$ 的表示法完全一致：

$$\overrightarrow{OA}\times\overrightarrow{OB}=\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{vmatrix}$$

$$=(a_2b_3-a_3b_2)\boldsymbol{i}+(a_3b_1-a_1b_3)\boldsymbol{j}+(a_1b_2-a_2b_1)\boldsymbol{k}.$$

只需把 $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$ 分别换成 $\boldsymbol{j}\wedge\boldsymbol{k},\boldsymbol{k}\wedge\boldsymbol{i},\boldsymbol{i}\wedge\boldsymbol{j}$ 即可. 同样，$\wedge_n^3(\mathbf{R}^n)$ 对应于 $3\times n$ 矩阵，它构成通常的 3 维空间，而过去学过的初等的立体几何在此都有效.

最值得注意的是 $\wedge_n^n(\mathbf{R}^n)$. $\dim\wedge_n^n(\mathbf{R}^n)=1=\dim\wedge_n^0(\mathbf{R}^n)$. $\wedge_n^0(\mathbf{R}^n)$ 中之元是实数. $\wedge_n^n(\mathbf{R}^n)$ 之元是什么？上面已经说了 $\wedge_n^n(\mathbf{R}^n)$ 是 $n$ 阶反对称协变张量之空间. 取 $\mathbf{R}^n$ 之任一基底 $\{e_1,\cdots,e_n\}$，则其对偶基 $\{e^1,\cdots,e^n\}$ 是 $(\mathbf{R}^n)^*$ 之一个基底，而 $e^1\wedge\cdots\wedge e^n$ 是 $\wedge^n(\mathbf{R}^n)$ 的基底. 张量成分则是 $\omega_{1,2,\cdots,n}$，只要看一下它们在 $\mathbf{R}^n$ 的基底即标架的变换下如何变换就可以知道 $\Lambda^n(\mathbf{R}^n)$ 中的元究竟是什么了. 于是设在 $\mathbf{R}^n$ 中作基底的变换 (1)：

$$f_i=a_i^je_j,\ \{f_i\}=A\{e_j\}.$$

于是 $\{e_i\}$ 之对偶基将按逆步变换变成 $\{f_i\}$ 之对偶基

$$\{e^i\}=a_j^i\{f^j\},\ \{e^i\}={}^tA\{f^j\}.$$

以下我们记逆步矩阵 ${}^tA^{-1}=B=(b_i^j)$. 于是有

**定理 7**　在上述变换下

$$\tilde{\omega}_{1,2,\cdots,n} = \det(a_i^j)\,\omega_{1,2,\cdots,n}, \tag{29}$$

$$f^1 \wedge \cdots \wedge f^n = \det(b_i^j)\, e^1 \wedge \cdots \wedge e^n. \tag{30}$$

这里 $\tilde{\omega}_{1,2,\cdots,n}$ 是张量成分在基底 $\{f_i\}$ 下的表示.

**证**　因为 $\tilde{\omega}_{1,2,\cdots,n}$ 是协变张量成分,故

$$\begin{aligned}\tilde{\omega}_{1,2,\cdots,n} &= a_1^{j_1}\cdots a_n^{j_n}\omega_{j_1,\cdots,j_n}\\ &= \sum_\sigma \operatorname{sgn}\sigma\, a_1^{\sigma(1)}\cdots a_n^{\sigma(n)}\omega_{1,2,\cdots,n}\\ &= \det(a_i^j)\,\omega_{1,2,\cdots,n}.\end{aligned}$$

这里 $\sigma$ 是变 $(1,2,\cdots,n)$ 为 $(j_1,j_2,\cdots,j_n)$ 之排列.

其次,由 $f^i = b_j^i e^j$ 有

$$\begin{aligned}f^1 \wedge \cdots \wedge f^n &= b_{j_1}^1\cdots b_{j_n}^n e^{j_1} \wedge \cdots \wedge e^{j_n}\\ &= \sum_\sigma \operatorname{sgn}\sigma b_{\sigma(1)}^1\cdots b_{\sigma(n)}^n \cdot e^1 \wedge \cdots \wedge e^n\\ &= \det(b_i^j)\, e^1 \wedge \cdots \wedge e^n.\end{aligned}$$

前面我们已多次讲到空间 $\mathbf{R}^n$ 的定向.其实空间 $\mathbf{R}^n$ 的定向即给定一个基底的次序,例如 $\{e_1,\cdots,e_n\}$.对偶基 $\{\omega^1,\cdots,\omega^n\}$ 的相应次序,既可以认为是给出了 $(\mathbf{R}^n)^*$ 之定向也可认为是给出了 $\mathbf{R}^n$ 的定向.所以若改变 $\{\omega^i\}$ 之次序为例如 $\{\omega^2,\omega^1,\omega^3,\cdots,\omega^n\}$,也可认为 $\mathbf{R}^n$ 得到了相反的定向.但若对基底作一个变换,新基底的哪一种次序与给定次序的原基底所定义的定向一致?因此定义定向更好的办法是利用定理 7:设有 $\mathbf{R}^n$ 的两个基底 $\{e_i\}$ 与 $\{f_i\}$,自然存在一个非奇异的线性变换 $A = (a_j^i)$ 使 (1) 成立,因此可认为 $\{f_i\}$ 是 $\{e_i\}$ 的变换.若相应的 $\omega$ 变为 $\tilde{\omega}$,当有 (29) 成立.这里 $\det(a_i^j) \neq 0$.我们规定,若 $\det(a_i^j) > 0\,(<0)$,就说 $\{e_i\}$ 与 $\{f_i\}$ 规定相同(相反)的定向.于是 $\mathbf{R}^n$ 有两个定向,分别对应于 (29) 中行列式 $\det(a_i^j)$ 之符号正负.

至今我们还没有在 $\mathbf{R}^n$ 中引入度量.度量引入外代数以后会得到什么?我们所需的度量(包括欧氏度量以及十分重要的洛伦兹度量),基础在于定义两个向量的内积 $\langle u,v\rangle$.对内积有两个要求:1.双线性:即对 $u,v$ 分别均为线性的;2.对称性:$\langle u,v\rangle = \langle v,u\rangle$;3.本应加上正定性,但是我们刻意地要推广这一点.设有 $\mathbf{R}^n$ 的基底 $\{e_i\}$ 使 $u = u^i e_i$,$v = v^j e_j$,则记

$$\langle u,v\rangle = g_{ij}u^i v^j,\quad g_{ij} = \langle e_i,e_j\rangle. \tag{31}$$

通常讲到内积时都规定 $(g_{ij})$ 是正定矩阵,我们要放弃这一点,其原因在于物理学发展到麦克斯韦的电磁场理论以后就很清楚:必须把时间与空间统一起来看成一个 4 维时空连续统 $\{(x^0,x^1,x^2,x^3)\}$,$x^0 = ct$,$c$ 是光速.它称为闵可夫斯基时空(Minkowski space-time)记作 $M_4$,它是一个线性空间,当然也就是一个微分流形.但是在 $M_4$ 上不能使用欧几里得度量,因为这不能说明物理问题.我们需要的是如下的洛伦兹度量(Lorentz metric)

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2. \tag{32}$$

现在 $(g_{ij}) = \begin{pmatrix}1&&&\\&-1&&\\&&-1&\\&&&-1\end{pmatrix}$ 不是正定的,所以洛伦兹度量不是某种黎曼度量而时常说是

一种伪黎曼度量(pseudo-Riemannian metric).引入洛伦兹度量以研究 $M_4$ 是相对论的基础.相对论的出现是科学史上划时代的大事.所以我们宁可多一些麻烦也放弃 $(g_{ij})$ 正定的要求.代替它,我们要求相应于内积(30)的 $(g_{ij})$ 必须是非蜕化的:$\det(g_{ij}) \neq 0$.放弃正定性会带来一些麻烦.例如我们也讲o.n.系:对两个向量 $u, v$ 若 $\langle u, v\rangle = 0$ 就说 $u, v$ 正交,这一点和过去一样.若再能找一个实数 $c$,使 $cu$ 适合 $\langle cu, cu\rangle = c^2\langle u, u\rangle = 1$,则以 $cu = u/[\langle u, u\rangle]^{1/2}$ 代替 $u$ 将得到长为1的向量.在 $(g_{ij})$ 非正定时,可能 $\langle u, u\rangle < 0$(这在相对论中有明确的物理意义),从而 $c = \langle u, u\rangle^{-\frac{1}{2}}$ 成为虚数!因此 $u$ 不能规范化.所以现在o.n.系之定义中 $\langle u, u\rangle = 1$ 的条件将代以 $|\langle u, u\rangle| = 1$.尽管有了这样的麻烦,但一个最重要的性质仍得以保存:在 $\mathbf{R}^n$ 与 $(\mathbf{R}^n)^*$ 中可以建立同构 $\mathbf{R}^n \cong (\mathbf{R}^n)^*$ 如下:设有 $\mathbf{R}^n$ 的一个基底 $\{e_1, \cdots, e_n\}$ 与它在 $(\mathbf{R}^n)^*$ 中的对偶基 $\{\omega^1, \cdots, \omega^n\}$,对于 $\mathbf{R}^n \in u = u^j e_j$,我们要找 $(\mathbf{R}^n)^*$ 中的一个元 $u^* = u_j^* \omega^j$ 使对任意 $v \in \mathbf{R}^n$,$v = v^j e_j$,有

$$u^*(v) = \langle u, v\rangle = g_{ij}u^i v^j.$$

但是式左就是 $u_j^* v^j$,因此为了能建立 $u^*$ 与 $u$ 之一一对应,只需要

$$g_{ij}u^i = u_j^*$$

为一同构即可.为了使上式成为同构,必要充分条件是 $(g_{ij})$ 非蜕化,而不是正定.

为了要解上面的方程就应引入 $(g_{ij})$ 之逆矩阵 $(g^{ij})$.这里非常重要的是看到 $(g_{ij})$ 是一个二阶对称协变张量,而 $(g^{ij})$ 则是一个二阶对称逆变张量.我们可以用内积应与坐标的选择无关来证明 $(g_{ij})$ 是一个二阶对称协变张量.§1中我们已经就几个特例——如动能应该与坐标无关,$ds^2$ 应该与坐标无关等——看到了这一点.现在则一般地利用(31)式来证明它.于是设有 $\mathbf{R}^n$ 的两个基底 $\{e_1, \cdots, e_n\}$ 与 $\{f_1, \cdots, f_n\}$ 且 $f_i = a_i^j e_j$,这两个坐标系中度量是 $(g_{ij})$,$g_{ij} = \langle e_i, e_j\rangle$ 与 $(\tilde{g}_{ij})$,$\tilde{g}_{ij} = \langle f_i, f_j\rangle$.如果向量 $u, v$ 在这两个基底下的坐标分别为 $(u^i)$,$(v^i)$ 与 $(\tilde{u}^i)$,$(\tilde{v}_i^i)$,则由内积与坐标无关可得

$$\langle u, v\rangle = g_{ij}u^i v^j = \tilde{g}_{ij}\,\tilde{u}^i\,\tilde{v}^j = (\tilde{g}_{kl}a_i^k a_j^l)\,u^i v^j.$$

因为这个等式对任意 $(u^i)$,$(v^j)$ 都成立,故有

$$g_{ij} = \tilde{g}_{kl}a_i^k a_j^l.$$

而知 $(g_{ij})$ 是二阶协变张量.类似地,利用 $(g^{ij}) = (g_{ij})^{-1}$,即 $g^{ik}g_{kj} = \delta_j^i$ 可证明 $(g^{ij})$ 是二阶对称逆变张量.通过计算逆矩阵容易看到,对于洛伦兹度量,有 $g_{00} = g^{00} = 1$,$g_{ij} = g^{ij} = -\delta_{ij}$,$g^{0i} = g_{i0} = 0\ (i \neq 0)$.

$(g_{ij})$ 与 $(g^{ij})$ 还有一个重要作用,即可利用它们来实现协变张量与逆变张量的互换.例如,令 $\{e_1, \cdots, e_n\}$ 是 $\mathbf{R}^n$ 的一个基底,则令 $e^j = g^{ij}e_i$ 易证 $\{e^1, \cdots, e^n\}$ 是 $(\mathbf{R}^n)^*$ 的一个基底,特别重要的是,若 $\{e_1, \cdots, e_n\}$ 是一个关于 $(g_{ij})$ 的o.n.系,则 $\{e^1, \cdots, e^n\}$ 是其对偶基底,事实上

$$\langle e^i, e_j\rangle = g^{ij}\langle e_k, e_j\rangle = g^{ik}g_{kj} = \delta_j^i.$$

在§1中我们常将 $(\mathbf{R}^n)^*$ 之基底记为 $\{\omega^1, \cdots, \omega^n\}$,下面当 $\{e_1, \cdots, e_n\}$ 是o.n.系时我们不再用 $\{\omega^i\}$ 的记号而用 $\{e^i\}$.不过要注意,当基底 $\{e_i\}$ 按(1)变换时,$\{e^i\}$ 一般并不按逆步变换而变.

最重要的是要研究,当 $\mathbf{R}^n$ 中赋有非蜕化的度量 $(g_{ij})$ 时,$\wedge^n(\mathbf{R}^n)$ 的基底 $e^1 \wedge \cdots \wedge e^n$ 有何具体特征.首先我们要注意,若取 $\mathbf{R}^n$ 之基底 $\{e_1, \cdots, e_n\}$ 为关于 $(g_{ij})$ 的o.n.系,因为 $\langle e_i, e_j\rangle = \pm\delta_{ij}$,故由(31)

$$\det(g_{ij}) = \det(\langle e_i, e_j\rangle) = \pm 1.$$

现在令基底按(1)实行变换，度量$(g_{ij})$若变为$(\tilde{g}_{ij})$，则因$(g_{ij})$是二阶协变张量，故

$$\tilde{g}_{ij} = a_i^k a_j^l g_{kl},$$

因此

$$\tilde{g} = \det(\tilde{g}_{ij}) = [\det(a_i^k)]^2 \det(g_{ij}) = (\det A)^2 g.$$

如果我们限制变换(1)要保持定向不变，从而 $\det A > 0$，则当$\{f_1,\cdots,f_n\}$是o.n.系，从而$|\tilde{g}| = 1$时，

$$1 = \sqrt{|g|}(\det A), \sqrt{|g|} = \det(A^{-1}).$$

引用定理7之第二式，立即有

$$\begin{aligned} f^1 \wedge \cdots \wedge f^n &= \det(b_i^j) e^1 \wedge \cdots \wedge e^n \\ &= \sqrt{|g|} e^1 \wedge \cdots \wedge e^n. \end{aligned} \tag{33}$$

这是一个很重要的不变式. 它告诉我们不论基底$\{e_1,\cdots,e_n\}$如何取，只要以一个o.n.系的基底$\{f_1,\cdots,f_n\}$为准而且保持定向不变，(33)式右方之值不变. 把这些结果总结起来，我们有

**定义6**　设$\mathbf{R}^n$中有非蜕化的度量$(g_{ij})$，以某一个关于$(g_{ij})$的o.n.系基底为准，所有与它具有相同定向的基底$\{e_1,\cdots,e_n\}$，$\sqrt{|g|}e^1 \wedge \cdots \wedge e^n$是不变的，称为$\mathbf{R}^n$在此度量下的体积元(volume element).

**注**　对于$\mathbf{R}^n$的另一个定向，体积元自然是$-\sqrt{|g|}e^1 \wedge \cdots \wedge e^n$.

以上我们看到，若$\mathbf{R}^n$上有内积则必有非蜕化度量$(g_{ij})$. 反过来，有了$(g_{ij})$也一定有内积$\langle u, v\rangle$，而且适合(31)式. 这时，$(\mathbf{R}^n)^*$上也可以定义内积：若$u = u_i f^i$，$v = v_j f^j$则与(31)对偶，可以定义

$$\langle u, v\rangle = g^{ij} u_i v_j. \tag{34}$$

总之，有了度量$(g_{ij})$即也可在$\wedge^1(\mathbf{R}^n)$上定义内积. 从而产生了一个问题：可否在$\wedge^p(\mathbf{R}^n)$，$0 \leqslant p \leqslant n$上都定义内积？事实上任取两个$p$向量$\omega = \omega^1 \wedge \cdots \wedge \omega^p$；$\sigma = \sigma^1 \wedge \cdots \wedge \sigma^p$，因为$\wedge^p(\mathbf{R}^n)$中之元都是$p$向量之线性组合，所以只要能定义$\langle \omega, \sigma\rangle = \langle \omega^1 \wedge \cdots \wedge \omega^p, \sigma^1 \wedge \cdots \wedge \sigma^p\rangle$，就可以在$\wedge \Lambda^p(\mathbf{R}^n)$上定义内积.

**定理8**　若$\mathbf{R}^n$赋有非蜕化的度量$(g_{ij})$，则在$\wedge^p(\mathbf{R}^n)$，$0 \leqslant p \leqslant n$，上可以定义两个$p$向量$\omega, \sigma$之内积为

$$\langle \omega, \sigma\rangle = \langle \omega^1 \wedge \cdots \wedge \omega^p, \sigma^1 \wedge \cdots \wedge \sigma^p\rangle = \det(\langle \omega^i, \sigma^j\rangle). \tag{35}$$

**证**　(35)显然是(34)之推广. 为了证明它是内积，只要证明以下各点即可.

(i) $\langle \omega, \sigma\rangle = \langle \sigma, \omega\rangle$是自明的.

(ii) 多重线性. 因为$\det(\langle \omega^i, \sigma^j\rangle)$对于例如$\omega^i$或$\sigma^j$均为线性的，所以$\langle \omega, \sigma\rangle$对每一个$\omega^i$与$\sigma^j$，$1 \leqslant i, j \leqslant p$，均为线性的.

(iii) $\langle \lambda\omega, \sigma\rangle = \lambda\langle \omega, \sigma\rangle$是自明的.

(iv) 非蜕化性. 即应证明若对$\forall \sigma \in \wedge^p(\mathbf{R}^n)$均有$\langle \omega, \sigma\rangle = 0$，则$\omega = 0$. 为证明这一点，注意到$\omega^i, \sigma^j$均为$\wedge^1(\mathbf{R}^n) = (\mathbf{R}^n)^*$之元，在$(\mathbf{R}^n)^*$中相对于$(g_{ij})$取o.n.基底$\{e^1,\cdots,e^n\}$，我们只要证明$\{e^{i_1} \wedge \cdots \wedge e^{i_p}\}$，$i_1 < \cdots < i_p$，是$\wedge^p(\mathbf{R}^n)$之o.n.系即可. 如果证明到了这一点，则

$\{e^{i_1}\wedge\cdots\wedge e^{i_p}\}$ 是线性无关的，而因为 $\{\cdot\}$ 中含有 $\binom{n}{p}$ 个元，$\binom{n}{p}=\dim\wedge^p(\mathbf{R}^n)$，所以 $\{e^{i_1}\wedge\cdots\wedge e^{i_p}\}$ 就成了 $\wedge^p(\mathbf{R}^n)$ 之基底. 于是非蜕化性就容易证明了. $\{e^{i_1}\wedge\cdots\wedge e^{i_p}\}$ 是 o.n. 系，证如下：

$$\langle e^{i_1}\wedge\cdots\wedge e^{i_p}, e^{j_1}\wedge\cdots\wedge e^{j_p}\rangle=\det(\langle e^{i_s},e^{j_r}\rangle).$$

设 $i_1<i_2<\cdots<i_p$；$j_1<j_2<\cdots<j_p$. 如果 $j_1<i_1$，则 $j_1<$ 一切 $i_s$，$1\leqslant s\leqslant p$，从而 $\langle e^{i_s},e^{j_1}\rangle=0$，即是说上式右方行列式有一列元素全为 0，从而式右与式左均为 0，而只有在 $i_1=j_1$ 时，上式双方可以不为 0. 于是令 $i_1=j_1$，再看 $j_2$，这时已有 $j_2>j_1=i_1$，如果 $j_2<i_2<\cdots<i_p$，则 $\langle e^{i_s},e^{j_2}\rangle=0$. 因此又只有当 $i_2$ 也与 $j_2$ 相等式上式双方可以不为 0. 仿此进行下去

$$\langle e^{i_1}\wedge\cdots\wedge e^{i_p}, e^{j_1}\wedge\cdots\wedge e^{j_p}\rangle=\begin{cases}0 & (i_1,\cdots,i_p)\neq(j_1,\cdots,j_p),\\ \det(\langle e^{i_s},e^{i_r}\rangle)=\det(\delta_r^s), & (i_1,\cdots,i_p)=(j_1,\cdots,j_p).\end{cases}$$

但因 $\langle e^{i_s},e^{i_s}\rangle=\pm1$，故当 $(i_1,\cdots,i_p)=(j_1,\cdots,j_p)$ 时式右为一对角形行列式，其主对角线上之元为 $\pm1$，从而其值亦为 $\pm1$. 准确些说，其值为 $(-1)^s$，$s$ 是基底中适合 $\langle e^i,e^i\rangle=-1$ 的元素之个数，定理证毕.

上面我们已经看到，有了内积之后，就可以定义 $\mathbf{R}^n$ 与 $(\mathbf{R}^n)^*$ 之同构. 现在 $\wedge^p(\mathbf{R}^n)$ 也有了内积. 我们可以利用它来定义 $\wedge^p(\mathbf{R}^n)$ 与 $\wedge^{n-p}(\mathbf{R}^n)$ 的一个同构. 事实上，若 $\omega\in\wedge^{n-p}(\mathbf{R}^n)$，$\sigma\in\wedge^p(\mathbf{R}^n)$ 则 $\omega\wedge\sigma\in\wedge^n(\mathbf{R}^n)$，但 $\dim\wedge^n(\mathbf{R}^n)=1$，其基底只含一个元，即体积元 $e^1\wedge\cdots\wedge e^n$. 这里我们假设了 $\{e^1,\cdots,e^n\}$ 是此内积（度量）下的 o.n. 系基底，故

$$\omega\wedge\sigma=A(\omega,\sigma)e^1\wedge\cdots\wedge e^n.$$

这里 $A(\omega,\sigma)$ 是一个实数，而且因为 $e^1\wedge\cdots\wedge e^n$ 与 $\omega,\sigma$ 无关，易见 $A(\omega,\sigma)$ 对 $\omega$ 与 $\sigma$ 是双线性的. 如果固定 $\omega$，则 $A(\omega,\sigma)$ 是 $\sigma\in\wedge^p(\mathbf{R}^n)$ 之线性泛函. 但因 $\wedge^p(\mathbf{R}^n)$ 上已有了内积，故必存在 $\wedge^p(\mathbf{R}^n)$ 之一个元 $*\omega$（这个元显然线性地依赖于 $\omega$，故有一个线性算子 $*$，使此元可记为 $*\omega$），使 $A(\omega,\sigma)=\langle*\omega,\sigma\rangle$，于是我们有

**定理 9**　上面定义的算子 $*$ 称为霍奇(Hodge) $*$ 算子，它是 $\wedge^{n-p}(\mathbf{R}^n)$ 到 $\wedge^p(\mathbf{R}^n)$ 上的同构：

$$\omega\wedge\sigma=\langle*\omega,\sigma\rangle e^1\wedge\cdots\wedge e^n. \tag{36}$$

**证**　$*:\wedge^{n-p}(\mathbf{R}^n)\to\wedge^p(\mathbf{R}^n)$ 是线性映射是明显的. 现证它是单射. 为此设有某个 $\omega$ 使 $*\omega=0$. 于是对任意 $\sigma\in\wedge^p(\mathbf{R}^n)$ 有 $\omega\wedge\sigma=0$. 因此 $\omega=0$ 而 $*$ 是单射. 但 $\wedge^{n-p}(\mathbf{R}^n)$ 与 $\wedge^p(\mathbf{R}^n)$ 是维数同为 $\binom{n}{p}$ 的线性空间，所以只要 $*$ 是一个单射线性映射，则必为同构，定理证毕.

**注**　我们在构造 $*$ 算子时使用了一个特定的 o.n. 系 $\{e_1,\cdots,e_n\}$. 实际上 $*$ 之定义与 $\{e_i\}$ 之选取无关. 这里我们略去证明，而始终应用 $\mathbf{R}^n$ 关于度量 $(g_{ij})$ 下的 o.n. 系 $\{e^1,\cdots,e^n\}$：$e^i=g^{ij}e_j$.

现在我们在这个基底下计算 $*$ 算子的具体表达式. 为此，只需对 $\wedge^p(\mathbf{R}^n)$ 的基底 $\{e^{i_1}\wedge\cdots\wedge e^{i_p}\}$ 中的任意元 $\omega=e^{i_1}\wedge\cdots\wedge e^{i_p}$ 来计算 $*\omega$. 由定义

$$(e^{i_1}\wedge\cdots\wedge e^{i_p})\wedge\sigma=\langle*\omega,\sigma\rangle e^1\wedge\cdots\wedge e^n,\ \forall\sigma\in\wedge^{n-p}(\mathbf{R}^n). \tag{37}$$

这里不妨设 $i_1<i_2<\cdots<i_p$. 令 $\{j_1,\cdots,j_{n-p}\}$ 是 $\{i_1,\cdots,i_p\}$ 在 $\{1,2,\cdots,n\}$ 中的余集，而且令 $j_1<\cdots<j_{n-p}$. 于是除非 $\sigma=he^{j_1}\wedge\cdots\wedge e^{j_{n-p}}$（$h\neq0$ 为一实数），(37) 之左方必为 0. 这样就

知道 $*\omega$ 可用 $\wedge^{n-p}(\mathbf{R}^n)$ 之基底 $\{e^{j_1}\wedge\cdots\wedge e^{j_{n-p}}\}$ 来表示，即存在非零实数 $c$，使

$$*\omega = ce^{j_1}\wedge\cdots\wedge e^{j_{n-p}},\ (c\neq 0, c\in\mathbf{R}), \tag{38}$$

而 (37) 成为

$$he^{i_1}\wedge\cdots\wedge e^{i_p}\wedge e^{j_1}\wedge\cdots\wedge e^{j_{n-p}} = \frac{h}{c}\langle e^{j_1},e^{j_1}\rangle\cdots\langle e^{j_{n-p}},e^{j_{n-p}}\rangle e^1\wedge\cdots\wedge e^n.$$

令 $\pi$ 为变 $\{i_1,\cdots,i_p;j_1,\cdots,j_{n-p}\}$ 为 $\{1,2,\cdots,n\}$ 之排列，由上式消去 $h$ 即有

$$\frac{1}{c} = \operatorname{sgn}\pi\langle e^{j_1},e^{j_1}\rangle\cdots\langle e^{j_{n-p}},e^{j_{n-p}}\rangle.$$

因为 $\langle e^{j_1},e^{j_1}\rangle = g^{j_1j_1}$ 而 $(g^{ij}) = (g_{ij})^{-1}$ 为对角矩阵，$g^{j_1j_1} = 1/g_{j_1j_1} = \pm 1$，所以

$$c = \pm 1.$$

代入 (38) 即可得到 $*(e^{i_1}\wedge\cdots\wedge e^{i_p})$ 的表达式.

$*$ 算子有下面的重要性质.

**定理 10** 霍奇 $*$ 算子具有以下性质：

(1) 对于 $\omega\in\wedge^p(\mathbf{R}^n)$

$$**\omega = (-1)^{p(n-p)+s}\omega, \tag{39}$$

$s$ 是 o.n. 基底 $\{e_1,\cdots,e_n\}$ 中适合条件 $\langle e^j,e^j\rangle = -1$ 的元素之个数.

(2) 若 $\omega,\sigma\in\wedge^p(\mathbf{R}^n)$，则

(a) $\omega\wedge*\sigma = \sigma\wedge*\omega = (-1)^s\langle\omega,\sigma\rangle e^1\wedge\cdots\wedge e^n$.

(b) $\langle*\omega,*\sigma\rangle = (-1)^s\langle\omega,\sigma\rangle$.

**证** (1) 我们只需对 $\lambda = e^{i_1}\wedge\cdots\wedge e^{i_p}\ (i_1<\cdots<i_p)$ 证明 (39) 即可. 令 $\{j_1,\cdots,j_{n-p}\}$ 是 $\{i_1,\cdots,i_p\}$ 在 $\{1,2,\cdots,n\}$ 中的余集，而且 $j_1<\cdots<j_{n-p}$，由 $*$ 算子的定义 (38) 有

$$\begin{aligned}**\omega = \operatorname{sgn}\rho\cdot\operatorname{sgn}\pi\langle e^{i_1},e^{i_1}\rangle\cdots\langle e^{i_p},e^{i_p}\rangle\langle e^{j_1},e^{j_1}\rangle\cdots\\ \langle e^{j_{n-p}},e^{j_{n-p}}\rangle\omega = (-1)^s\operatorname{sgn}\rho\cdot\operatorname{sgn}\pi\,\omega\end{aligned} \tag{40}$$

$\rho$ 与 $\pi$ 分别是把 $\{1,2,\cdots,n\}$ 变为 $\{i_1,\cdots,i_p,j_1,\cdots,j_{n-p}\}$ 与 $\{j_1,\cdots,j_{n-p},i_1,\cdots,i_p\}$ 之置换. 但是，为将例如 $\{i_1,\cdots,i_p,j_1,\cdots,j_{n-p}\}$ 对换成为 $\{i_1,\cdots,i_{p-1},j_1,\cdots,j_{n-p},i_p\}$ 共需 $n-p$ 个对换；再把 $i_{p-1}$ 与 $j_1,\cdots,j_{n-p}$ 对换，仿此进行下去，总之要有 $p(n-p)$ 个对换. 所以 $\operatorname{sgn}\rho = (-1)^{p(n-p)}\operatorname{sgn}\pi$. 代入 (40) 即得 (39) 式.

(2) (a) 由外积之反交换性（定理 4）以及 $*$ 算子的定义有

$$\begin{aligned}\omega\wedge*\sigma &= (-1)^{p(n-p)}*\sigma\wedge\omega = (-1)^{p(n-p)}\langle**\sigma,\omega\rangle e^1\wedge\cdots\wedge e^n\\ &= (-1)^s\langle\sigma,\omega\rangle e^1\wedge\cdots\wedge e^n = (-1)^s\langle\omega,\sigma\rangle e^1\wedge\cdots\wedge e^n\\ &= \sigma\wedge*\omega.\end{aligned}$$

(b) 我们有

$$\omega\wedge*\sigma = \langle*\omega,*\sigma\rangle e^1\wedge\cdots\wedge e^n.$$

与 (a) 比较，即有

$$\langle*\omega,*\sigma\rangle = (-1)^s\langle\omega,\sigma\rangle.$$

现在我们要用霍奇 $*$ 算子来解释 $\mathbf{R}^3$ 中古典的向量分析公式. 我们仍用箭头或黑体字母作为古典的向量分析公式的向量记号. 于是 $\boldsymbol{i} = e^1, \boldsymbol{j} = e^2, \boldsymbol{k} = e^3$. 在 $\mathbf{R}^3$ 中使用笛卡儿直角坐标

系，使得协变向量与逆变向量不分，这样 $e^i = e_i$，而我们不必担心上下指标的区别. 这时 $n = 3$，$s = 0$. 设有三个向量：$\boldsymbol{A} = a_1\boldsymbol{i} + a_2\boldsymbol{j} + a_3\boldsymbol{k} = a_1e_1 + a_2e_2 + a_3e_3$，$\boldsymbol{B}$，$\boldsymbol{C}$ 的记法与此类似. 于是考虑 $p = 1$ 的情况，注意到这时 $p(n-p) = 2$，很容易看到会出现指标的轮换：

$$*\boldsymbol{i} = *e_1 = e_2 \wedge e_3,\ *\boldsymbol{j} = *e_2 = e_3 \wedge e_1,\ *\boldsymbol{k} = *e_3 = e_1 \wedge e_2.$$

$p = 2$ 时 $n - p = 1$，$p(n-p) = 2$，所以情况会是一样的，利用定理 10 的 (1) 有

$$*(e_2 \wedge e_3) = **e_1 = e_1,\quad *(e_3 \wedge e_1) = e_2,\quad *(e_1 \wedge e_2) = e_3.$$

$p = 0$ 与 $p = 3$ 的情况又是一样的. 于是

$$*1 = e_1 \wedge e_2 \wedge e_3,\ *(e_1 \wedge e_2 \wedge e_3) = 1.$$

利用这些公式我们可以很容易地解释古典的向量分析中的向量积与混合积. 对于前者，我们有

$$\boldsymbol{A} \times \boldsymbol{B} = (a_2b_3 - a_3b_2)\boldsymbol{i} + (a_3b_1 - a_1b_3)\boldsymbol{j} + (a_1b_2 - a_2b_1)\boldsymbol{k}.$$

现在则求 $\boldsymbol{A} \wedge \boldsymbol{B}$ 得

$$\boldsymbol{A} \wedge \boldsymbol{B} = (a_2b_3 - a_3b_2)e_2 \wedge e_3 + (a_3b_1 - a_1b_3)e_3 \wedge e_1 + (a_1b_2 - a_2b_1)e_1 \wedge e_2.$$

于是

$$\boldsymbol{A} \times \boldsymbol{B} = *(\boldsymbol{A} \wedge \boldsymbol{B}). \tag{41}$$

在讨论混合积之前还要讲一下内积(即标量积). 前面已说过 $*$ 算子定义的基础是内积，反过来内积也可以用 $*$ 来表示. 实际上

$$\begin{aligned} *\boldsymbol{A} \wedge \boldsymbol{B} &= (a_1e_2 \wedge e_3 + a_2e_3 \wedge e_1 + a_3e_1 \wedge e_2) \wedge (b_1e_1 + b_2e_2 + b_3e_3) \\ &= (a_1b_1 + a_2b_2 + a_3b_3)e_1 \wedge e_2 \wedge e_3, \end{aligned}$$

因此

$$\boldsymbol{A} \cdot \boldsymbol{B} = *(*\boldsymbol{A} \wedge \boldsymbol{B}) = a_1b_1 + a_2b_2 + a_3b_3. \tag{42}$$

在古典的向量分析中，以 $\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}$ 棱的平行六面体之体积用一个非常奇怪的公式来表示：

$$V = \boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B} \cdot (\boldsymbol{C} \times \boldsymbol{A}) = \boldsymbol{C} \cdot (\boldsymbol{A} \times \boldsymbol{B}).$$

但用 $*$ 算子则非常清楚. 只要用分量来表示，利用 (42) 与 (41) 即有

$$\begin{aligned} \boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) &= *[\boldsymbol{A} \wedge *(\boldsymbol{B} \times \boldsymbol{C})] = *(\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C}). \\ &= \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \cdot 1 = V. \end{aligned}$$

这里 $V$ 是上述平行六面体之体积. 总之把 1 写成 $*(e_1 \wedge e_2 \wedge e_3)$ 即知

$$V = *\left(\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} e_1 \wedge e_2 \wedge e_3\right) = *(\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C}). \tag{43}$$

这里我们应用了定理 7 中的 (30).

总之，利用了霍奇 $*$ 算子，立即有

$$\boldsymbol{A} \times \boldsymbol{B} = \text{以 } \boldsymbol{A}, \boldsymbol{B} \text{ 为边的平面四边形的有符号的面积} = *(\boldsymbol{A} \wedge \boldsymbol{B})$$

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \text{以 } \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \text{ 为棱的平行六面体的有符号的体积} = *(\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C}).$$

这就很明确地体现了格拉斯曼把几何量分成互相有联系的不同类型的量之系统这个思想.

把格拉斯曼的这个思想用到微分流形上将得到更丰硕的结果.

# §4　外微分形式

**1. 从一个张量空间到张量丛**　设 $M$ 是一个 $n$ 维 $C^\infty$ 微分流形，于是对其上每一点 $P\in M$，有切空间 $T_P(M)$ 与余切空间 $T_P^*(M)$ 是其对偶. 如果孤立地看一个点，即固定 $P$，则切空间与余切空间就是 $\mathbf{R}^n$ 与 $(\mathbf{R}^n)^*$，其上的向量与张量结构我们已在 §1 与 §3 中作了详尽的讨论. 若令 $P$ 在 $M$ 中变动，我们就会得到一族线性空间 $\bigcup\limits_P T_P(M)$ 与 $\bigcup\limits_P T_P^*(M)$. 因为 $M$ 是光滑的，我们直觉地会想到，例如 $T_P(M)$ 中某一切向量当 $P$ 变动时也会光滑地变化. 但是实际上并不如此简单. 不同的 $P$ 点的切空间之基底互相没有关系，当 $P$ 一旦变动，原来 $T_P(M)$ 的基底(例如记作 $\{e_{1P},\cdots,e_{nP}\}$) 不知道会怎样变. 甚至原来线性无关的向量也会变成线性相关而使 $\{e_{1P},\cdots,e_{nP}\}$ 不再成为基底. 所以在 §2 中讨论向量丛时，我们不能只考虑底空间 $M$ 上之点 $P$ 附近 $U_\alpha$ 中的局部坐标映射，还要考虑纤维之变化，于是要求 $\pi^{-1}(U_\alpha)$ 能同胚于 $U_\alpha\times\mathbf{R}^n$，且 $\pi^{-1}(P)\simeq\mathbf{R}^n$. 这样一来，当 $P$ 在 $U_\alpha$ 中变化时，可以使 $P$ 点在每个位置上的基底之间有线性同构关系，且此同构之系数是 $C^\infty(M)$ 函数. 这样一来，底空间的局部坐标 $x$ 与纤维之坐标 $\xi$ 就完全分开，而在 $\bigcup\limits_{P\in U_\alpha} T_P(M)$ 中可以引入如下的局部坐标系 $(x_1,\cdots,x_n;\xi_1,\cdots,\xi_n)$. 这一点正是向量丛定义中的要点，即所谓局部平凡化条件. 有了这一点，我们上面提出的困难就自然消解了. 但又会产生一个新问题：原来固定 $P$ 点讨论 $T_P(M)$ 时是把它作为一个线性空间，其系数均为实数，现在"系数"要随 $P$ 变动了，自然就应取为 $C^\infty(U_\alpha)$ 函数. 看起来这只是毫不起眼的小事，实际上带来很大的问题：线性空间的系数构成一个域(field). 在代数学中域是有特定含义的，简单地说就是在域中可以作四则运算，唯有以零作分母除外. 现在 $C^\infty(U_\alpha)$ 只是一个环(ring)，而一个环简单说来就是可以在其中作加、减、乘的集合，而除法一般是不许可的. 设若有一个不恒为 0 的 $C^\infty$ 函数 $f(x)$. $f(x)$ 当然不是零元，但它可能在某一点 $P$ 为 0：$f(P)=0$，于是对这个非 0 元也不能作 $1/f(x)$，即 $1/f(x)\notin C^\infty$. 这是一个不小的限制. 例如元素为实数(这是我们一直在用的域)的矩阵 $A$，是可以讨论其逆矩阵的(尽管当 $\det A=0$ 时逆矩阵并不存在)，但元素在某一环(例如 $C^\infty(U_\alpha)$) 中的矩阵，就根本不能谈它的逆矩阵问题，因为逆矩阵的元是以 $\det A$ 为分母的. 读者一定要记住，元素在环中的矩阵和元素在域中的矩阵全然不同. 例如，对于前者就不会有很有用的若尔当标准形. §2 中已说过，这种系数为 $C^\infty$ 函数的"线性空间"称为一个模(module). 模和线性空间大不相同. 在讨论向量丛时应该注意这一点. 但是我们又不打算去讨论模的问题，所以对读者们说来，最简单的办法是不要轻易地做除法，而要特别注意分母是否在某点可能为 0！以上我们重述了关于丛的问题.

现在即以 $T_P(M)$ 为基本的线性空间，即上节中的 $\mathbf{R}^n$，或 $V$. 而在每一点 $P\in M$ 可以作一个张量代数 $\mathscr{T}_P$，一个格拉斯曼代数 $\wedge_P(M)$，于是我们有张量代数丛 $\mathscr{T}(M)$ 和格拉斯曼代数丛(或外代数丛) $\wedge(M)$. 在 §2 中我们定义了一个向量丛的切口的概念. 设有 $M$ 上的向量丛 $E$，其相应的投影是 $\pi$，若有自 $U\subset M$ 到 $E$ 中的映射 $\sigma$，适合 $\pi\circ\sigma=\mathrm{id}$ 就说 $\sigma$ 是 $E$ 在 $U$ 上的切口，若 $\sigma$ 是光滑的，就说此切口是光滑的. 如果用比较直观的语言，现以切丛为例，并设 $U$ 是一个坐标邻

域，其局部坐标为$(x)$，所谓给它一个切口，就是对$U$中每一点(记为$x$)，可以在$E$中$x$处的纤维中找到一个元与之对应. 因为在$U$中各点处的纤维都同构于$\mathbf{R}^n$，而且有一个与$x$无关的，即适用于整个$U$上的基底$\{e\}$，所以纤维中的元若用$\{e\}$表示必为一个特定的$(\xi_1(x),\cdots,\xi_n(x))\simeq\sum_{i=1}^{n}\xi_i(x)e_i$. 所谓光滑切口则是指所有这些$\xi_i(x)\in C^\infty(U)$. 这样我们得到坐标邻域$U$中的切口，就是一个局部切口. 是否有可能由局部切口转而构造出$M$上的整体切口?我们当然会想到找到两个相交的坐标邻域$U$与$V$以及其中的局部坐标$\{x\}$与$\{y\}$，于是在$U\cap V$中上述特定的纤维中的元$(\xi_1(x),\cdots,\xi_n(x))$或写为$(\eta_1(y),\cdots,\eta_n(y))$之间可以用迁移函数的雅可比矩阵来转换. 但是这样做是否可以在整个$M$上作出一个切口是一个严重的问题. 不但这种整体问题有如此的困难，有时甚至局部问题也有困难. 我们趁这个机会向读者介绍一件值得注意的事. 能不能找到适用于坐标邻域$U$中的o.n.系?o.n.系显然来自内积，而内积则决定了一个度量. 在一个局部坐标系下，由每一点处的一个二阶对称协变张量决定. 如果这个张量当$x$在$U$中变化时，是张量丛的$C^\infty$切口，因此用矩阵$(g_{ij}(x))$表示，每一个$g_{ij}(x)\in C^\infty(U)$，且为简单设计它是正定的. 要找一个o.n.系就是找一个坐标变换$x=x(y)$，使在$(y)$坐标下$(g_{ij})$成为一对角矩阵$\lambda(x)I$，即要求

$$\tilde{g}_{kl}(y)=g_{ij}(x)\frac{\partial x^i}{\partial y^k}\frac{\partial x^i}{\partial y^l}=0,\quad k\neq l$$
$$\tilde{g}_{11}(y)=\cdots=\tilde{g}_{nn}(y)=\lambda(x).$$

$\lambda(x)$之值不定. 由对称性，上式实际上是含$\left(\frac{n}{2}+1\right)(n-1)$个方程的方程组，而我们想由它决定出$n$个函数$x_i=x_i(y_1,\cdots,y_n)$来，我们知道$\left(\frac{n}{2}+1\right)(n-1)\geqslant n\ (n\geqslant 2)$，而且除了$n=2$外，上式中有严格的“>”成立. 所以方程个数多于未知函数个数，而这个方程组是超定的(over-determined)因此一般是不可解的. 首先指出这一点的正是黎曼本人，而且就见于他著名的就职演说(habilitation)“论作为几何基础的假设”中. 按这个说法，对维数大于2的微分流形，一般在哪怕任意小的坐标邻域中，在切空间上也不一定有o.n.系(但在2维流形上却可以)，所以下面我们能讲某一点切空间中的o.n.系而不能讲切丛的某一切口中的o.n.系.

于是回到主题. 对$n$维微分流形$M$，我们就以$T_P(M)$作为上节中的$\mathbf{R}^n$. 于是在一个坐标邻域$U$中，相对于局部坐标$(x^i)$ $T_P(M)$有标准的基底$\left\{\frac{\partial}{\partial x^1},\cdots,\frac{\partial}{\partial x^n}\right\}$，$T_P^*(M)$则有标准的对偶基底$\{\mathrm{d}x^1,\cdots,\mathrm{d}x^n\}$. 这里的$\mathrm{d}x^i$是余切向量而不是坐标，它是$\mathrm{d}f$当$f(x)\equiv x^i$时的特例. 若作用在切向量$u=(u^1,\cdots,u^n)$上，$\mathrm{d}x^i(u)=u^i$，即它可以把$u$的第$i$个分量$u^i$(它是一个实数)“取”出来. 如果把切向量写成微分算子$X$，则由§2中的定义$\mathrm{d}f(X)=X(f)$，于是$(\mathrm{d}x^i)(X)=X(x^i)$，把$X$写成$X=u^i\frac{\partial}{\partial x^i}$则$(\mathrm{d}x^i)(X)=X(x^i)=u^i$. 这与利用对偶性$\langle\mathrm{d}x^i,\frac{\partial}{\partial x^j}\rangle=\delta_j^i$得出的结果是一致的.

利用局部坐标$(x^i)$，余切向量丛的$C^\infty$切口成为$\omega(x)=a_i(x)\mathrm{d}x^i$，$a_i(x)\in C^\infty(U)$，称为1微分形式. 切向量丛的$C^\infty$切口是$X=a^i(x)\frac{\partial}{\partial x^i}$，$a^i(x)\in C^\infty(U)$称为向量场. $U$上的向量

场集合记为$\mathscr{X}(U)$.$\mathscr{T}^s_r(U)$的切口称为张量场,其局部坐标表示为$a^{j_1\cdots j_s}_{i_1\cdots i_r}(x)\dfrac{\partial}{\partial x^{j_1}}\otimes\cdots\otimes\dfrac{\partial}{\partial x^{j_s}}\otimes$ $\mathrm{d}x^{i_1}\otimes\cdots\otimes\mathrm{d}x^{i_r}$. $\wedge^r(U)$的切口成为

$$\omega(x)=a_{i_1\cdots i_r}(x)\mathrm{d}x^{r_1}\wedge\cdots\wedge\mathrm{d}x^{i_r},i_1<\cdots<i_r,\tag{1}$$

$a_{i_1\cdots i_r}(x)\in C^\infty(U)$称为一个$r$微分形式或$r$外微分形式.我们记$r$微分形式的集合为$F^r(U)$, $r=0,1,\cdots,n$. (1)中的指标时常为方便起见还加上限制$i_1<\cdots<i_r$.若不加这个限制,则$r$微分形式常写为

$$\frac{1}{r!}a_{i_i\cdots i_r}(x)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_r}.\tag{1'}$$

以下是否采用了这个限制将可从上下文中看到.

外代数中的代数运算法则都很容易转移到外微分形式中来.但应该注意由于实数域$\mathbf{R}$现在被光滑函数环$C^\infty$取代,所以有些公式会有很容易的变化.首先是,0微分形式现在不是实数域$\mathbf{R}$,而是$C^\infty$函数环.此外

$$f(\omega+\sigma)=f\omega+f\sigma$$

中$f$不只是实数,而是$C^\infty$函数,关于外积则有

$$\omega\wedge f\sigma=f\omega\wedge\sigma=f(\omega\wedge\sigma),$$

这里例如$f\sigma$变成0微分形式与另一微分形式之乘积.这时,0微分形式之"位置"可以随意移,这与反交换性

$$\omega\wedge\Lambda\sigma=(-1)^{rs}\sigma\wedge\omega$$

不矛盾,因为若$\sigma=f$,则$s=0$,上式自然成为

$$\omega\wedge f=f\wedge\omega=f\omega.$$

最后一步其实可以看成一个"定义",即0微分形式与$\omega$之"标量积"就是外积.

**2. 推前与拉回** 现在我们要讨论,若有微分流形$M$到$N(\dim M=m,\dim N=n)$的映射(我们恒限于可微映射)$\varphi$,则$\wedge(M)$与$\wedge(N)$之间是否有一个相应的映射?首先,$M$上的$P$点的切向量$X_P$将变为$N$上$\varphi_{(P)}$处的切向量$Y_{\varphi(P)}$,§2中指出

$$Y_{\varphi(P)}=\mathrm{d}_P\varphi\cdot X_P.\tag{2}$$

这里$\mathrm{d}_P\varphi$是切映射,在局部坐标$x,y$下,若把$\varphi$写成

$$y=y(x),y^i=y^i(x^1,\cdots,x^m),\tag{3}$$

则$\mathrm{d}_P\varphi$之坐标表示就是(3)在$P$点的雅可比矩阵

$$\mathrm{d}_P\varphi=\frac{\partial(y^1,y^2,\cdots,y^n)}{\partial(x^1,x^2,\cdots,x^m)}.\tag{4}$$

所以(2)中的$\mathrm{d}_P\varphi$既可以作与坐标无关的理解,像§2中指出的曲线的等价类$[c]$之间的映射或$T_P(M)$与$T_{\varphi(P)}(N)$的向量之间的映射一样是一个整体,一个完整的记号;也可以采用坐标表示(3)而看成为微分学一章中讲的,即映射(3)的线性部分,即泰勒展开式的线性部分.这时$\mathrm{d}_P\varphi$不必再看成一个抽象的完整不可分的记号,而它确实就是对古典的函数$\varphi$(即(3))作古典的微分运算(即(4)).想要转到向量丛上去,就简单地去掉下标中的$P$,就是让$P$变动.但这时$\mathrm{d}_P$变成了d却反而要"找"这个"躲起来"的$P$,于是找到了$P$的坐标$x$,而d就是对这个$x$作古典的求导运

算.§3 第一段末引用了 M.Spivak 一段话,说这些老记号都有了新定义,要我们“慢慢来跟”,下面我们还要去跟这个微分 d.

在 §2 中我们把 $\mathrm{d}\varphi$ 写成 $\varphi_*$,而由 (2) 以及 (4) 可得切映射 $\varphi_*: T_PM \to T_{\varphi(P)}N$ 的局部坐标表示

$$\mathrm{d}\varphi_P \cdot \left(\frac{\partial}{\partial x}\right) = \left.\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^m)}\right|_P \left(\frac{\partial}{\partial y}\right) = A\left(\frac{\partial}{\partial y}\right).$$

我们在 §2 中说 $\varphi_*$ 是一个推前,因为它映射的方向与 $\varphi: M \to N$ 是一致的.于是我们看见有一类与流形相联系的几何量在映射 $\varphi$ 下是被推前的.但还有一类对象是被拉回的.§2 中我们已见到一个例子,即函数.设函数 $f \in C^\infty(N)$ 是定义在 $N$ 上的.如果 $\varphi$ 把 $M$ 的一个区域映射到 $N$ 中 $f$ 之定义域内,则 $f\circ\varphi$ 定义在 $M$ 的这一个区域内.这里一定会产生一个问题:$M$ 中的这个区域不一定是整个 $M$,$\varphi$ 把它映到 $f$ 之定义域内也不一定是映到整个定义域上,更不一定映到整个 $N$ 上.此外,这个映射 $\varphi$ 一般不可能是微分同胚.至少从 $\dim M$ 与 $\dim N$ 不一定相同就可以看到这一点.对于这样一些麻烦我们现在不来讨论,我们关心的只是映射的“方向”.§2 的 (25) 式告诉我们(现将记号稍微改变一下).

$$(\mathrm{d}\varphi_P \cdot X)(f)_{\varphi(P)} = X(f\circ\varphi)_P.$$

我们把 $f\circ\varphi$ 记作 $\varphi^* f$,而说是将定义在 $N$ 上的 $f$ 拉回到 $M$ 上,并且仿照矩阵与内积的记号将上式写成

$$\langle \varphi_* X, f\rangle_{\varphi(P)} = \langle X, \varphi^* f\rangle_P. \tag{5}$$

于是看到 $\varphi_*$ 与 $\varphi^*$ 的关系很像一个矩阵与转置矩阵的关系,而 (5) 式很像内积.如果我们让 $P$ 在 $M$ 上变动(至少是在 $M$ 的一个区域中变动),则应略去 $\langle,\rangle$ 下的下标,而得到两个 $C^\infty$ 函数,左方定义在 $N$ 上,右方定义在 $M$ 上.

上面我们看到,推前与拉回的区别恰好和逆变与协变的区别相应.这不是偶然的,因为二者均以对偶性为基础.为了把这一点说明白,我们不妨从最简单的情况看起.现在用 $\mathbf{R}^n$ 来代替 $T_PM$.在 §1 中讨论逆变与协变两个概念时,我们是以其在坐标变换——现在就假设是线性变换——下的的表现为依据的.于是不妨设 $\mathbf{R}^n$ 中已有了一个基底 $\{e_1,\cdots,e_n\}$ 并且经过变换成为 $\{f_1,\cdots,f_n\}$:

$$e_i = a_i^j f_j.$$

在 $(\mathbf{R}^n)^*$ 中则有对偶基 $\{\omega^1,\cdots,\omega^n\}$ 与 $\{\sigma^1,\cdots,\sigma^n\}$.于是 $\mathbf{R}^n$ 中的向量 $\alpha$ 以及 $(\mathbf{R}^n)^*$ 中的向量 $\beta$ 分别用

$$\alpha = x^i e_i, \beta = \xi_j \omega^j$$

表示.而 $\beta$ 作用在 $\alpha$ 上之值,即线性泛函 $\beta$ 在 $\alpha$ 上之值用“内积”表示:

$$\langle \beta, \alpha\rangle = \xi_i x^i.$$

如果换用新的基底 $\{f_i\}$ 以及相应的对偶基底 $\{\sigma^j\}$,则因例如 $\alpha$ 本身就是同一个向量,不应因坐标之选择而有异.所以 $\alpha$ 在新坐标系下的表示 $\alpha = y^j f_j$ 应该与原来的向量相同:

$$x^i e_i = y^j f_j.$$

但因 $e_i = a_j^i f_j$,代入上式有

$$y^j f_j = x^i a_i^j f_j,$$

所以

$$y^j = a_i^j x^i.$$

我们把它用矩阵写成

$$y = Ax.$$

同样的理由用于$(\mathbf{R}^n)^*$，就得到同一个向量$\beta$在两个坐标系下的表示$\beta=\xi_i x^i$与$\beta=\eta_j\sigma^j$中应有关系式

$$\xi = B\eta.$$

为什么写成$\xi=B\eta$而不写成$\eta=B\xi$，下面的演算自会说明.

问题在于$B$是什么？要回答这个问题，就要注意到，不但$\alpha,\beta$不应因坐标的改变而异，它们的内积$\langle\beta,\alpha\rangle$也是这样.所以

$$\langle\beta,\alpha\rangle = \xi_i x^i = \eta_j y^j.$$

但是$y^j=a_i^j x^i$，代入上式就有

$$\eta_j a_i^j x^i = \xi_i x^i.$$

因为此式应该对一切$x^i$都成立，所以

$$\xi_i = a_i^j \eta_j.$$

所以自然得到形如$\xi=B\eta$的关系式.而且注意到，现在是对$a_i^j$之上标求和.前面$y^j=a_i^j x^i$是对$a_i^j$之下标求和，所以立即有

$$B = {}^tA.$$

如果用矩阵形式来写，就更清楚：由于

$$\langle\eta, Ax\rangle = \langle{}^tA\eta, x\rangle = \langle\xi, x\rangle,$$

所以

$$\xi = B\eta = {}^tA\eta.$$

在§1中我们讲的是如何从“老坐标”变为“新坐标”，所以上式要改写为

$$\eta = {}^tA^{-1}\xi.$$

逆步变换即由此而来.

上面我们把§1的内容重复了一次.现在我们要换一个观点来看待它.上面我们把$T_PM$换成了$\mathbf{R}^n$，然后按§1的讲法在$\mathbf{R}^n$中作坐标变换.但是$\mathbf{R}^n$中换用了另一个坐标系以后，也可以看成是另一个$n$维线性空间.所以，如果把$T_PM$看成原来的$\mathbf{R}^n$，则现在的$\mathbf{R}^n$就可以认为是$T_QN$. $Q=\varphi(P)$，而$A$就认为是切映射$\mathrm{d}_P\varphi=\varphi_*: T_PM\to T_QN$.我们得到的则是其对偶映射：$B=\varphi^*$.逆步变换${}^tA^{-1}$中采用转置是因为我们从$\mathbf{R}^n$转到$(\mathbf{R}^n)^*$.其中出现“逆”，则表明了$\varphi^*$的映射方向与$\varphi_*$相反：一是拉回，一是推前.

上面讲的当然不完满，这一点前面已经说过了：至少$\dim M$与$\dim N$不一定相同.和前面一样，我们不去讨论这些问题.至少我们看到，协变逆变也好，推前拉回也好，其基础就是对偶性.我们把它写成(5)式，借用了线性代数的记号，正是因为其基本思想是一致的.

于是我们看到了0阶协变张量(即一阶微分形式)之拉回有了定义，一阶协变张量(即一阶微分形式)之拉回也有了定义，那么$\mathscr{T}_r^0(M)$乃至$\Lambda^r(M)$之切口$T$之拉回$\varphi^*T$又应如何定义呢？

我们不问更一般的张量如何处理，因为在几何学与物理学中很少见到高阶逆变张量．注意到高阶协变张量可以用 0 阶与 1 阶协变张量，用$\otimes$与 $Alt$ 来生成，所以只要弄清$\otimes$与 $\Lambda$ 与 $\varphi^*$ 之关系，则一切迎刃而解．这就是下面定义 1 与定理 1，2 之来由，以及它们与张量的协变、逆变性质之关系．

今设 $T$ 为$N$上一个$r$阶协变张量．$X_1,\cdots,X_r$ 为$M$上$r$个向量场．现用 $\varphi_*$ 把它们推前成为$N$上的$r$个向量场：$\varphi_* X_1,\cdots,\varphi_* X_r$，于是 $T$ 可以作用于它们：$T(\varphi(x))(\varphi_* X_1,\cdots,\varphi_* X_r)$，回到 $M$ 来看这就是对$M$上的$r$个向量场给出了一个多重线性形式，即对固定的 $x\in M$，

$$(X_1(x),\cdots,X_r(x))\rightarrow \mathbf{R},$$

它自然是 $x$ 点的 $\mathscr{T}_r^0$ 丛的一个元．当 $x$ 在$M$上变动时，则得到 $\mathscr{T}_r^0(M)$ 丛的一个 $C^\infty$ 切口，即 $M$ 上一个张量场．我们就用它作为 $T$ 的拉回的定义：

**定义 1**　设 $T$ 是$N$上一个$r$阶协变张量场，$r\geqslant 0$，当 $r>0$ 时，对 $M$ 上任意$r$个向量场 $X_1,\cdots,X_r$ 可以得到一个$C^\infty$ 函数：$T(x)(\varphi_* X_1,\cdots,\varphi_* X_r)$，它定义了 $M$ 上一个张量场，称为 $T$ 在 $M$ 上的拉回，于是得到拉回算子 $\varphi^*$ 之定义如下：

$$(\varphi^* T)(x)(X_1,\cdots,X_r) = T(\varphi(x))(\varphi_* X_1,\cdots,\varphi_* X_r).\tag{6}$$

$r=0$ 时，零阶张量即一个 $C^\infty$ 函数 $f$ 之拉回定义为

$$\varphi^* f = f\circ\varphi.\tag{7}$$

拉回算子 $\varphi^*$ 显然是一个线性算子．

(6) 式也可以模仿 (5) 式那样用对偶关系来表示：我们常把 (6) 写为

$$\langle\varphi^* T:X_1,\cdots,X_r\rangle = \langle T:\varphi_* X_1,\cdots,\varphi_* X_r\rangle.\tag{8}$$

它与 (5) 的区别在于 (5) 对 $f$ 是线性的，而 (8) 对 $X_1,\cdots,X_r$ 是多重线性的．但是如果只考虑一阶张量，即 $r=1$，则 (8) 告诉我们，$\varphi^*$ 与 $\varphi_*$ 分别作用在切空间的切口上与余切空间的切口上．而且 $\varphi^*$ 与 $\varphi_*$ 互相对偶．下面讲到 $\varphi^*$ 与 $\varphi_*$ 的局部坐标表示时这一点就表现得更清楚．此外要注意，(5) 表示一个微分算子对一个函数的作用，即一个 1 阶逆变向量作用于 0 阶协变张量，但是我们不知道什么是 0 阶逆变张量，所以也谈不上 1 阶协变张量作用于 0 阶逆变张量的问题．所以 (5) 与 (8) 其实并不一样，(8) 则是利用张量作为一个多重线性泛函之定义而全不涉及微分运算．

总之，$\varphi^*$ 现在已定义到张量代数上了．这个代数中有乘法运算 —— 张量积，于是我们当然要考虑拉回算子与张量积的关系．我们的结论是：二者是可交换的，即有

**定理 1**　设 $M$ 与$N$分别为$m$维与$n$维的微分流形，$\varphi:M\rightarrow N$ 为一可微映射．若 $\omega$ 与$\sigma$分别是 $N$ 上的$r$阶与$s$阶协变张量场，则

$$\varphi^*(\omega\otimes\sigma) = (\varphi^*\omega)\otimes(\varphi^*\sigma).\tag{9}$$

**证**　先固定一点 $P\in M$，并在 $T_PM$ 中任取$r+s$个向量场$X_i$ 使$X_i(p)\in T_PM$，于是

$$\begin{aligned}\varphi^*(\omega\otimes\sigma)(X_1,\cdots,X_{r+s})_P &= (\omega\otimes\sigma)(\varphi_* X_1,\cdots,\varphi_* X_{r+s})_{\varphi(P)}\\ &= \omega(\varphi_* X_1,\cdots,\varphi_* X_r)_{\varphi(P)}\cdot\sigma(\varphi_* X_{r+1},\cdots,\varphi_* X_{r+s})_{\varphi(P)}\\ &= (\varphi^*\omega)(X_1,\cdots,X_r)_P\cdot(\varphi^*\sigma)(X_{r+1},\cdots,X_{r+s})_P\\ &= (\varphi^*\omega\otimes\varphi^*\sigma)(X_1,\cdots,X_{r+s})_P.\end{aligned}$$

让 $P$ 在$M$中变动即得关于张量场的公式 (9)．

外积是张量积生成的，即对张量积再作用以交代算子，见§3的定义5(23)式

$$T \wedge S = \frac{(r+s)!}{r!s!} Alt(T \otimes S).$$

所以，现在要讨论 $\varphi^*$ 与 $Alt$ 是否可交换.这时有

**定理2**　$M, N$ 和 $\varphi$ 与定理1相同，若 $\omega$ 与 $\sigma$ 分别为 $N$ 上的 $r$ 微分形式与 $s$ 微分形式，则

$$\varphi^*(\omega \wedge \sigma) = (\varphi^*\omega) \wedge (\varphi^*\sigma). \tag{10}$$

**证**　我们先证 $\varphi^*$ 与 $Alt$ 可换.事实上，令 $P$ 为 $M$ 上一点，$X_i \in T_P(M)$，$i = 1,2,\cdots,r$，$\pi$ 是 $\{1,2,\cdots,r\}$ 的任一排列.对于 $N$ 上任一个 $r$ 阶协变张量场 $T$，由定义1，

$$(\varphi^*T)_P(X_{\pi(1)},\cdots,X_{\pi(r)}) = T(\varphi_*X_{\pi(1)},\cdots,\varphi_*X_{\pi(r)})_{\varphi(P)},$$

所以

$$\sum_\pi \operatorname{sgn}\pi \cdot (\varphi^*T)_P(X_{\pi(1)},\cdots,X_{\pi(r)}) = \sum_\pi \operatorname{sgn}\pi \cdot T(\varphi_*X_{\pi(1)},\cdots,\varphi_*X_{\pi(r)})_{\varphi(P)}.$$

由 $Alt$ 之定义，再让 $P$ 在 $M$ 中变动，有

$$Alt(\varphi^*T)(X_1,\cdots,X_r) = \varphi^*(Alt\ T)(X_1,\cdots,X_r).$$

这就是我们需要的可换性：

$$Alt(\varphi^*\ \bullet) = \varphi^*(Alt\ \bullet). \tag{11}$$

把这个结果用于 $\omega$ 与 $\sigma$，并应用定理1，有

$$\begin{aligned}\varphi^*(\omega \wedge \sigma) &= \frac{(r+s)!}{r!s!}\varphi^*[Alt(\omega \otimes \sigma)] \\ &= \frac{(r+s)!}{r!s!}Alt[\varphi^*(\omega \otimes \sigma)] = \frac{(r+s)!}{r!s!}Alt(\varphi^*\omega \otimes \varphi^*\sigma) \\ &= \varphi^*\omega \wedge \varphi^*\sigma.\end{aligned}$$

定理证毕.

§2中我们得出了推前算子 $\varphi_*$ 的局部坐标表达式.现在要问 $\varphi^*$ 的局部坐标表达式是什么？为此，我们再把§2中相关的记号在此重复一次.令 $x = (x^1,\cdots,x^m)$ 和 $y = (y^1,\cdots,y^n)$ 分别是 $M$ 与 $N$ 中的局部坐标，而可微映射 $\varphi: M \to N$ 成为

$$y^i = y^i(x^1,\cdots,x^m),\ i = 1,2,\cdots,n.$$

在§2中得到，$\varphi_*$ 的表达式如下：

$$\varphi_*\left(\frac{\partial}{\partial x^i}\right) = \mathrm{d}\varphi\left(\frac{\partial}{\partial x^i}\right) = \frac{\partial y^k}{\partial x^i} \cdot \frac{\partial}{\partial y^k}, \tag{12}$$

或

$$\varphi_*\left(\frac{\partial}{\partial x}\right) = \mathrm{d}\varphi\left(\frac{\partial}{\partial x}\right) = {}^tA\left(\frac{\partial}{\partial y}\right). \tag{13}$$

这里 $A = \left(\frac{\partial y^k}{\partial x^i}\right) = \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \cdots & \frac{\partial y^1}{\partial x^m} \\ & & \\ \frac{\partial y^n}{\partial x^1} & \cdots & \frac{\partial y^n}{\partial x^m} \end{pmatrix}$，而求和是对 $y$ 的上标进行的，所以上式中用记号 ${}^tA$.从这里我们看到 $\mathrm{d}\varphi$ 一方面应该看作是一个整体记号，而由切映射的与坐标无关的定义得出上式.另

一方面又可以看作是用通常微积分教本中讲的古典的微分运算作用于函数 $y^k = y^k(x^1,\cdots,x^n)$ 的结果.尽管其含义不同,计算结果都是一样的.现在我们同样利用这两种方法来计算 $\varphi^*$.

首先利用 $\varphi^*$ 的定义(8)式,并且只看 $T$ 为1阶协变张量场的情况,因为高阶张量场可以由定理1和2化为1阶张量场之张量积或外积.又因 $\varphi^*$ 是线性算子,所以只需设 $T$ 为1阶张量场之基底中的元 $\mathrm{d}y^k$(注意 $T$ 是 $N$ 上的张量)即可.因此我们只要计算 $\langle\varphi^*\mathrm{d}y^k, X\rangle$ 即可,而且不妨取 $X$ 即为 $M$ 上切向量的基底;$X = \dfrac{\partial}{\partial x^i}$.应用(5)与(12)有

$$\left\langle\varphi^*\mathrm{d}y^k,\frac{\partial}{\partial x^i}\right\rangle = \left\langle\mathrm{d}y^k,\varphi_*\left(\frac{\partial}{\partial x^i}\right)\right\rangle = \left\langle\mathrm{d}y^k,\frac{\partial y^l}{\partial x^i}\frac{\partial}{\partial y^l}\right\rangle = \frac{\partial y^k}{\partial x^i}.$$

因此

$$\varphi^*(\mathrm{d}y^k) = \frac{\partial y^k}{\partial x^i}\mathrm{d}x^i = A(\mathrm{d}x^i)\text{ 的第 }k\text{ 个元}. \tag{14}$$

它与(13)的区别在于:首先它是对 $x^i$ 求和即是对下标(上标表示所在行的位置,下标表示所在列的位置)求和因此其系数矩阵是(13)中系数矩阵之转置.其次,(13)式是把 $N$ 上的对象经 ${}^tA$ 变成 $M$ 上的对象之推前;(14)则是把 $M$ 上的 $\mathrm{d}x$ 经 $A$ 变成 $N$ 上对象 $\mathrm{d}y$ 之拉回.因此,(13)与(14)是互为逆步的.但我们没有用 ${}^tA^{-1}$ 这样的记号,因为一般说来 $m\neq n$,而 $A$ 不是方阵,自然无法作 $A^{-1}$.

我们再从另一个观点来看 $\varphi^*(\mathrm{d}y^k)$.§2中已说过用微分记号来表示切映射是因为切映射实际上确实就是古典的微分.$\mathrm{d}y$ 作为 $\dfrac{\partial}{\partial y}$ 的对偶基底正是以 $\mathrm{d}y$ 把 $\dfrac{\partial}{\partial y}$ 映入 $\mathbf{R}$ 的线性映射或线性泛函.所以 $\mathrm{d}y$ 正是把 $N$ 之切向量 $\dfrac{\partial}{\partial y}$ 映为 $\mathbf{R}$ 的切向量(仍旧是 $\mathbf{R}$ 中的元)的切映射,所以实质上也就是古典的微分.这一点在§2中就详细讨论过.把 $\mathrm{d}y$ 看作 $N$ 上的余切向量,就是把 $y$ 作为 $N$ 上的"自变量"得到的微分,$\varphi^*(\mathrm{d}y)$ 则是把它拉回到 $M$ 上,即是要把 $y$ 看成 $M$ 上的 $x$ 之函数,并问这时 $\mathrm{d}y$ 是什么.当然很清楚,由古典的微分学有

$$\mathrm{d}y^k = \frac{\partial y^k}{\partial x^i}\mathrm{d}x^i.$$

所以 $\varphi^*\mathrm{d}y$ 与 $\mathrm{d}y$ 之区别就是前者把 $y$ 看成 $x$ 的函数,而后者把 $y$ 看成自变量.这样我们又一次得到(14).这样的讲法虽然思路是明白的,但总有不严格之嫌.正确的作法应该是证明对任一 $C^\infty$ 函数 $f$,

$$\varphi^*(\mathrm{d}f) = \mathrm{d}(\varphi^* f), \tag{15}$$

这样就会有 $\varphi^*\mathrm{d}y^k = \mathrm{d}(\varphi^* y^k) = \mathrm{d}y^k(x)$ 而又得(14).(15)是一个下面就要讲的重要结果之特例.对于 $N$ 上的 $C^\infty$ 函数 $f$,因为 $\mathrm{d}f\in T^*N$,用 $\varphi^*$ 拉回后有 $\varphi^*\mathrm{d}f\in T^*M$.任取 $M$ 的一个向量场 $X\in TM$,应有

$$\begin{aligned}\varphi^*\mathrm{d}f(X) &= \mathrm{d}f(\varphi_* X) = (\varphi_* X)(f) = X[f\circ\varphi]\\ &= X(\varphi^* f) = \mathrm{d}(\varphi^* f)(X).\end{aligned}$$

所以(15)成立.

总之 $\varphi_*$ 与 $\varphi^*$ 的局部坐标表示是互为逆步的雅可比矩阵.

$\varphi_*$ 与 $\varphi^*$ 之性质还有一个重要的区别,即它们的链式法则方向不同.设有三个微分流形及

其间的可微映射

$$M \xrightarrow{\varphi} N \xrightarrow{\psi} L.$$

于是 $\psi \circ \varphi: M \to L$ 是一个可微映射. 对于切映射 $(\psi \circ \varphi)_*$, §2 的定理 5 指出

$$(\psi \circ \varphi)_* = \psi_* \circ \varphi_*. \tag{16}$$

然而对于拉回映射，则会得到一个次序与此相反的关系式:

$$(\psi \circ \varphi)^* = \varphi^* \cdot \psi^*. \tag{17}$$

为证此式，回到定义 1. 令 $T$ 是 $L$ 上任一 $r$ 阶协变张量场，$X_1, \cdots, X_r$ 为 $M$ 上任意 $r$ 个向量场，于是由于 $\psi \circ \varphi: M \to L$ 是一个可微映射，故由(6)式

$$\begin{aligned}(\psi \circ \varphi)^* T(x)(X_1, \cdots, X_r) &= T(x)[(\psi \circ \varphi_*)X_1, \cdots, (\psi \circ \varphi)_* X_r] \\ &= T(x)[\psi_*(\varphi_* X_1), \cdots, \psi_*(\varphi_* X_r)] = (\psi^* T)(x)[\varphi_* X_1, \cdots, \varphi_* X_r] \\ &= [\varphi^*(\psi^* T)](x)(X_1, \cdots, X_r).\end{aligned}$$

而(17)得证.

不论是从局部坐标表示看还是从链式法则的形式看，推前与拉回的关系都是互相逆步的，有如向量之协变与逆变的关系一般. 所以在选用记号时，凡下加 $*$ 者都是指映射方向与基本的可微映射 $\varphi: M \to N$ 一致的，而上加 $*$ 者都是与它相反的.

**3. 微分流形的定向** $\mathbf{R}^n$ 的定向可以由其某一坐标 $(x^1, \cdots, x^n)$ 的次序决定. 如果 $\mathbf{R}^n$ 还有另一个坐标 $(y^1, \cdots, y^n)$，则 $y^i = a^i_j x^j$，而 $a^i_j = \dfrac{\partial y^i}{\partial x^j}$，从而上述变换式之系数矩阵 $A = (a^i_j) = \dfrac{\partial(y^1, \cdots, y^n)}{\partial(x^1, \cdots, x^n)}$. 如果 $\det\left(\dfrac{\partial(y^1, \cdots, y^n)}{\partial(x^1, \cdots, x^n)}\right) > 0$，就说这两个坐标决定相同定向，若此行列式为负，则说它们定出相反定向. 这个概念转到微分流形 $M$ 上时会遇到新问题. 设 $M$ 可以用坐标邻域 $\{(U, \varphi)\}$ 来覆盖，而 $\varphi(U) \subset \mathbf{R}^n$ 中有局部坐标例如 $(x^1, \cdots, x^n)$，则我们可以认为它决定了 $U$ 中的定向. 所以在 $M$ 上局部给出定向是没有困难的，若换另一个坐标邻域 $V$ 而 $V \cap U \neq \varnothing$，我们要看 $V$ 中的局部坐标 $(y^1, \cdots, y^n)$ 与 $(x^1, \cdots, x^n)$ 在 $U \cap V$ 中是否决定相同定向. 如果是，就说 $(y^1, \cdots, y^n)$ 在 $V$ 中(而不只是在 $U \cap V$)中也决定与 $U$ 相同的定向. 这样，定向适用的区域就从 $U$ 向 $U$ 以外拓展. 我们应该限制只看 $M$ 为连通的情况，因为 $M$ 若不是连通的，而由若干个连通分支合成: $M = M_1 \cup M_2 \cup \cdots$，则应逐一地讨论 $M_i$ 中的定向. 对于连通的 $M$, $M = \bigcup_i U_i$，每个 $U_i$ 中均有局部坐标 $(x^1_{(i)}, \cdots, x^n_{(i)})$，而且每个 $U_i$ 均会与至少一个 $U_j$, $j \neq i$ 相交(除非 $M$ 就是由一个坐标邻域构成的: $M = U_1$)，而我们得到一系列局部坐标. 如果每当 $U_i \cap U_j \neq \varnothing$ 时恒有 $\det\left(\dfrac{\partial(x^1_{(i)}, \cdots, x^n_{(i)})}{\partial(x^1_{(j)}, \cdots, x^n_{(j)})}\right) > 0$，则由它们在 $U_i$ 与 $U_j$ 中分别决定的定向在 $U_i \cap U_j$ 中是相同的. 这时我们说 $U_i$ 与 $U_j$ 有相同定向，并以它为 $U_i \cup U_j$ 中的定向. 用这样的方法，如果每个 $U_i$ 中均有相同的定向，我们即以它作为 $M$ 之定向而说微分流形 $M$ 是可定向的. 如要在 $M$ 上找不到具有这样性质的 $\{U_1, \cdots, U_N, \cdots\}$ (不一定限于可数多个). 就说 $M$ 不可定向. 如果一个流形是可定向的，则由 $(U, \varphi)$: $(x^1, x^2, \cdots, x^n)$ 开始得到一个定向，而由 $(U, \varphi)$: $(x^2, x^1, \cdots, x^n)$ 开始就可以得到另一个定向. 这个概念和前面讲的曲面的两侧是相应的. 本书只讨论可定向的流形. 不可定向流形最为人熟知的是默比乌斯带和偶数维的射影空间 $RP^{2k}$, $k \geqslant 1$.

正如 $\mathbf{R}^n$ 的定向可以用 $\wedge^n(\mathbf{R}^n)$ 中指定的一个元来表示一样，对于微分流形 $M$，我们可以证明 $M$ 可定向的充分必要条件是在 $M$ 上存在一个处处非零的 $n$ 微分形式 $\Omega = f(x)\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$. 事实上，若两个局部坐标邻域 $U_1 \cap U_2 \neq \varnothing$，而在 $U_1, U_2$ 中分别有局部坐标 $(x^1, \cdots, x^n)$ 与 $(y^1, \cdots, y^n)$，$\Omega$ 和 $\tilde{\Omega}$ 则分别表示为 $f(x)\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n$ 和 $\tilde{f}(y)\mathrm{d}y^1 \wedge \cdots \wedge \mathrm{d}y^n$. 于是在 $U_1 \cap U_2$ 中应有

$$\begin{aligned} f(x)\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n &= \tilde{f}\,\mathrm{d}y^1 \wedge \cdots \wedge \mathrm{d}y^n \\ &= \tilde{f}(y)\det\left(\frac{\partial(y^1, \cdots, y^n)}{\partial(x^1, \cdots, x^n)}\right)\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n, \end{aligned}$$

这里与 $f(x)$ 与 $\tilde{f}(y)$ 均不为 0. 必要时交换 $x^1$ 与 $x^2$ 或 $y^1$ 与 $y^2$ 可以设 $f(x) > 0$，$\tilde{f}(y) > 0$，而由上式即得

$$\det\left(\frac{\partial(y^1, \cdots, y^n)}{\partial(x^1, \cdots, x^n)}\right) > 0.$$

这就符合原来 $M$ 可定向的要求. 反过来，若在 $M$ 上可以找到这样一个图册 $\{(U_i, \varphi_i)\}$，使相应的局部坐标适合上式，亦必可以找到一个在 $M$ 上处处不为 0 的 $n$ 微分形式.

**4. 外微分**　在 $M$ 上的 $C^\infty$ 微分形式代数中有一些很重要的微分运算. 其中最重要者当推外微分运算. 它以许多在数学物理中最常见的运算如梯度、散度、旋度为其特例. 现在我们详细地讨论外微分运算 —— 即外微分算子 d.

设 $f \in F^0(M)$，即 $f$ 为 $M$ 上的 $C^\infty$ 函数. 在 §2 中定义了 $f$ 的微分 $\mathrm{d}f$ 是向量场空间上的线性泛函：

$$\mathrm{d}f(X) = X(f).$$

因此 $\mathrm{d}f \in F^1(M)$. 如果用局部坐标表表示则有

$$\mathrm{d}f = \frac{\partial f}{\partial x^i}\mathrm{d}x^i,$$

$\mathrm{d}x^i$ 是 $x^i$ 作为坐标函数：$M \supset U \to \mathbf{R}$ 的微分，其实也就是古典的微分，这些都在 §2 中详细讨论过了. 现在我们要把它推广到 $F^p(M)$ 上来.

**定理 3**　设 $M$ 为一 $n$ 维微分流形，则必存在唯一的算子

$$\mathrm{d}: F^p(M) \to F^{p+1}(M),$$

称为 $M$ 上的外微分，具有以下性质：

(i) d 是 $R$ 线性的，而且具有所谓的"反导子"(anti-derivation) 性质：

$$\mathrm{d}(\omega \wedge \theta) = \mathrm{d}\omega \wedge \theta + (-1)^p\omega \wedge \mathrm{d}\theta, \omega \in F^p(M); \tag{18}$$

(ii) d 在 $F^0(M)$ 上就是通常的微分算子.

(iii) $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$.

**证**　先证唯一性. 为此，取 $M$ 的一个坐标邻域即区图 $(U, \varphi)$，并记其局部坐标为 $(x^1, \cdots, x^n)$. 于是任意 $\omega \in F^p(M)$ 必可在 $U$ 中表示为

$$\omega(x) = a_{i_1 \cdots i_p}(x)\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_p},$$

于是利用 d 的 $R$ 线性 (即对实数系数的线性) 以及 (18) 式，有

$$\mathrm{d}\omega(x) = \mathrm{d}a_{i_1 \cdots i_p} \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_p} + (-1)^0 a_{i_1 \cdots i_p}\mathrm{d}[\mathrm{d}x^{x_1} \wedge \cdots \wedge \mathrm{d}x^{i_p}].$$

这里利用了 $a_{i_1\cdots i_p}(x)\in F^0(M)$，又上式第二项为0.因为反复利用(18)，在 $\mathrm{d}[\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}]$ 的每一项中均出现 $\mathrm{d}^2x^i$ 形状之项，从而由(iii)必为0.总之，在 $M$ 之任一个区图中 $\mathrm{d}\omega$ 必只有唯一的表达式

$$\mathrm{d}\omega=\mathrm{d}a_{i_1\cdots i_p}\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}=\frac{\partial a_{i_1\cdots i_p}}{\partial x^j}\mathrm{d}x^j\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}.\tag{19}$$

所以 d 的唯一性得证.

存在性证明如下.在任一区图上用(19)来定义 $\mathrm{d}\omega$，我们现证这样的 d 具有所需的性质(i)～(iii).

先看(i)，d 是 $R$ 线性的是自明的.为证明它适合(18)，再取 $\theta=b_{j_1\cdots j_q}\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_q}$，于是有

$$\begin{aligned}\mathrm{d}(\omega\wedge\theta)&=\mathrm{d}[a_{i_1\cdots i_p}b_{j_1\cdots j_q}\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}\wedge\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_q}]\\&=(\mathrm{d}a_{i_1\cdots i_p}\cdot b_{j_1\cdots j_q}+a_{i_1\cdots i_p}\mathrm{d}b_{i_1\cdots j_q})\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{j_q}\\&=[\mathrm{d}a_{i_1\cdots i_p}\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}]\wedge[b_{j_1\cdots j_q}\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_q}]\\&\quad+(-1)^p[a_{i_1\cdots i_p}\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}]\wedge[\mathrm{d}b_{j_1\cdots j_q}\wedge\mathrm{d}x^{j_1}\wedge\cdots\wedge\mathrm{d}x^{j_q}]\\&=\mathrm{d}\omega\wedge\theta+(-1)^p\omega\wedge\mathrm{d}\theta.\end{aligned}$$

这里我们反复利用了(19)以及 $\mathrm{d}b_{j_1\cdots j_q}$ 是1微分形式，所以在将它移到 $\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$ 后面去时需要添加因子 $(-1)^p$.

(ii) 可由(19)直接得出.

(iii) 的证明可再用一次(19)，于是

$$\mathrm{d}\circ\mathrm{d}\omega=\mathrm{d}^2\omega=\frac{\partial^2a_{i_1\cdots i_p}}{\partial x^k\partial x^j}\mathrm{d}x^k\wedge\mathrm{d}x^j\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$$

把前面两个因子按 $k<j$ 的次序排列将有

$$\mathrm{d}^2\omega=\left(\frac{\partial^2a_{i_1\cdots i_p}}{\partial x^k\partial x^j}-\frac{\partial^2a_{i_1\cdots i_p}}{\partial x^j\partial x^k}\right)\mathrm{d}x^k\wedge\mathrm{d}x^j\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}=0,\qquad k<j.$$

余下的问题是：在两个区图 $U\cap V\neq\varnothing$时，应用 $U,V$ 中的局部坐标$(x^i)$，$(y^i)$分别按(19)式计算 $\mathrm{d}\omega$，结果是否一致?在 $U\cap V$ 中 $\omega$ 又可写为

$$\omega(x(y))=a_{i_1\cdots i_p}(x(y))\mathrm{d}x^{i_1}(y)\wedge\cdots\wedge\mathrm{d}x^{i_p}(y),$$

这里 $\mathrm{d}x^{i_1}(y)$ 就是前面讲的 $\varphi^*\mathrm{d}x^{i_1}$.如果用 $\mathrm{d}'$ 表示对局部坐标 $y$ 按(19)来作的外微分计算，则

$$\begin{aligned}\mathrm{d}'\omega&=\mathrm{d}'[a_{i_1\cdots i_p}(x(y))\mathrm{d}x^{i_1}(y)\wedge\cdots\wedge\mathrm{d}x^{i_p}(y)]\\&=\mathrm{d}'a_{i_1\cdots i_p}(x(y))\wedge\mathrm{d}x^{i_1}(y)\wedge\mathrm{d}x^{i_p}(y)+\cdots,\end{aligned}$$

"…"中的每一项都有 $\mathrm{d}'(\mathrm{d}x^i(y))$ 这样的因子.在这里我们已对 $a_{i_1\cdots i_p}$ 应用了一次反导子性质(18)，其中 $a_{i_1\cdots i_p}\in F^0(M)$.这个性质已经得到了证明.

但是 §2 中就已指出，$\mathrm{d}x^{i_1}(y)$ 就是函数 $x^{i_1}(y)$ 的通常的微分，所以

$$\mathrm{d}x^{i_1}(y)=\frac{\partial x^{i_1}(y)}{\partial y^j}\mathrm{d}y^j.$$

仿照上面证明(3)成立之法即有 $d'(dx^{i_1}(y))=0$. 又因为(19)指出,对于0微分形式外微分就是普通的微分,因此可以利用一阶微分的形式不变性:

$$d'a_{i_1\cdots i_p}(x(y))=da_{i_1\cdots i_p}(x).$$

把 $d'(dx^{x_1})=0$ 以及上式都代回 $d'\omega$ 之表达式,有

$$d'\omega=d\omega.$$

**注** 定理3的陈述中我们特别强调了 $R$ 线性. 其实这里没有新概念,只是说若取系数为实数——实常数——时它是线性的. 强调这一点是因为当我们从 $M$ 上一个固定点 $P$ 处的外代数过渡到 $P$ 在 $M$ 上变动从而得到一个丛时,系数将变成 $C^\infty$ 函数. 一个 $C^\infty$ 函数当然不能如同常数一样在微分运算 d 中"走进走出".

以上我们是用坐标表示来给出外微分 d 的定义,如果要给出一个与坐标无关的定义,则应看到,当 $\omega\in F^r(M)$ 时,$d\omega\in F^{r+1}(M)$,因此应该在给定 $r+1$ 个向量场 $X_1,\cdots,X_{r+1}$ 时求出 $d\omega(X_1,\cdots,X_{r+1})$ 之值并以此作为 $d\omega$ 之定义. 事实上,在给出了 $d\omega$ 之坐标表示后,我们可以证明下面的重要定理,而且可以用它作为 $d\omega$ 之定义. 这个定理中用到了一个重要概念,即两个一阶微分算子 $X$ 的交换子 $[X,Y]=X\circ Y-Y\circ X$ 的概念. 交换子概念无论在数学或物理中都极为重要,由于本书篇幅限制,不能详细讨论. 我们只能在 $X=a^i(x)\dfrac{\partial}{\partial x^i}$, $Y=b^i(x)\dfrac{\partial}{\partial x^i}$ 的情况下给出它在坐标系 $(x^1,\cdots,x^n)$ 下的表达式:

$$\begin{aligned}[X,Y]&=X\circ Y-Y\circ X=\left(a^i(x)\frac{\partial}{\partial x^i}\right)\left(b^j(x)\frac{\partial}{\partial x^j}\right)-\left(b^i(x)\frac{\partial}{\partial x^i}\right)\left(a^j(x)\frac{\partial}{\partial x^j}\right)\\&=\left(a^i\frac{\partial b^j}{\partial x^i}-b^i\frac{\partial a^j}{\partial x^i}\right)\frac{\partial}{\partial x^j}.\end{aligned}\tag{20}$$

所以 $[X,Y]$ 仍是一个一阶微分算子即向量场. 现在我们可以给出下面的与坐标无关的表达式

$$\begin{aligned}d\omega(X_1,\cdots,X_{r+1})&=\sum_{i=1}^{r+1}(-1)^{i+1}X_i(\omega(X_1,\cdots,\hat{X}_i,\cdots,X_{r+1}))\\&\quad+\sum_{1\leqslant i\leqslant j\leqslant r+1}(-1)^{i+j}\omega([X_i,X_j],X_1,\cdots,\hat{X}_i,\cdots,\hat{X}_j,\cdots,X_{r+1}),\end{aligned}\tag{21}$$

这里 $\hat{X}_i$ 表示 $X_i$ 已删去不出现. 这个公式计算比较复杂,而且以后我们又常只在 $p=1$ 时应用它. 所以下面只证明它的这个特例.

**定理4** 设 $\omega$ 是 $M$ 上的一个1微分形式,则对任意向量场 $X,Y$ 我们有

$$d\omega(X,Y)=X\cdot\omega(Y)-Y\cdot\omega(X)-\omega([X,Y]).\tag{22}$$

这里 $X\cdot\omega(Y)$ 表示 $X$ 作为一个算子作用在 $C^\infty$ 函数 $\omega(Y)$ 上. $Y\cdot\omega(X)$ 也相似.

**证** 如果用局部坐标表示 $\omega=a_i dx^i$, $X=X^i\dfrac{\partial}{\partial x^i}$, $Y=Y^i\dfrac{\partial}{\partial x^i}$,我们有

$$d\omega=da_i\wedge dx^i=\sum_{i<j}\left(\frac{\partial a_i}{\partial x^j}-\frac{\partial a_j}{\partial x_j}\right)dx^i\wedge dx^j,$$

$$\omega(X)=a_iX^i,\ \omega(Y)=a_iY^i,\ d\omega(X,Y)=\sum_{i,j}\left(\frac{\partial a_i}{\partial x^j}-\frac{\partial a_j}{\partial x^i}\right)X^iY^j.$$

注意,最后一个 $\sum$ 中 $i$ 不一定小于 $j$. 另一方面

$$X \cdot \omega(Y) = X(a_i Y^i) = \mathrm{d}(a_i Y^i)X = \frac{\partial a_i}{\partial x^j}Y^i X^j + a_i \frac{\partial Y^i}{\partial x^j}X^j,$$

$$Y \cdot \omega(X) = Y(a_i X^i) = \mathrm{d}(a_i X^i)Y = \frac{\partial a_i}{\partial x^j}X^i Y^j + a_i \frac{\partial X^i}{\partial x^j}Y^j.$$

这里我们应用了对偶基关系 $\mathrm{d}x^i\left(\frac{\partial}{\partial x^i}\right) = \delta^i_j$ 以及 $\mathrm{d}f \cdot X = X(f)$. 由交换子之定义(20)，有

$$\omega([X,Y]) = a_i\left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}\right).$$

综合这些式子即有

$$X \cdot \omega(Y) - Y \cdot \omega(X) - \omega([X,Y]) = \mathrm{d}\omega(X,Y).$$

下面的定理告诉我们外微分 d 与拉回算子是可交换的.

**定理 5** 对任意 $r$ 微分形式 $\omega$，

$$\mathrm{d}(\varphi^*\omega) = \varphi^*(\mathrm{d}\omega). \tag{23}$$

**证** $r = 0$ 时，此式即(15)式已经证明，所以下面只对 $r > 0$ 时求证. 为此我们用归纳法，设对 $(r-1)$ 微分形式 $\omega$ 已经证明了(23)，于是对 $\omega \in F^r(M)$，

$$\begin{aligned}\mathrm{d}(\varphi^*\omega) &= \mathrm{d}[\varphi^*(a_{i_1\cdots i_r}\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_{r-1}}) \wedge \varphi^*\mathrm{d}x^{i_r}]\\ &= \mathrm{d}[\varphi^*(a_{i_1\cdots i_r}\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_{r-1}})] \wedge \varphi^*\mathrm{d}x^{i_r} + (-1)^{r-1}\mathrm{d}[\varphi^*\mathrm{d}x^{i_r}]\\ &= \varphi^*(\mathrm{d}a \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_{r-1}}) \wedge \varphi^*(\mathrm{d}x^{i_r}) \qquad (\text{为简单起见略去了 } a \text{ 之下标})\\ &= \varphi^*(\mathrm{d}a \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_{r-1}} \wedge \mathrm{d}x^{i_r}) = \varphi^*(\mathrm{d}\omega).\end{aligned}$$

上面我们应用了归纳假设 $\mathrm{d}[\varphi^*(a\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^{i_{r-1}})] = \varphi^*\mathrm{d}[a\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_{r-1}}] = \varphi^*(\mathrm{d}a \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_{r-1}})$ 以及(15)式

$$\mathrm{d}[\varphi^*(\mathrm{d}x^{i_r})] = \mathrm{d}[\mathrm{d}(\varphi^* x^{i_r})] = \mathrm{d}^2(\varphi^* x^{i_r}) = 0.$$

外微分形式之所以重要，至少其原因之一在于，物理学中有重要应用的许多有关向量和张量的概念与公式都可以或应该用外微分形式来解释. 下面讲一点历史. 物理学中的许多向量引起数学家的重视大约是 19 世纪中叶的事. 为了研究这些量，人们曾提出了种种概念和名词. 前面我们提到了格拉斯曼的贡献. 尽管他的写作十分晦涩，很难为物理学家和工程师接受，但是他确曾考虑了他的理论对力学、磁学和晶体学的应用. 在这个方向最重要的贡献应该归功于哈密顿(W. R. Hamilton). 他的研究的出发点主要是物理学. 为了找出作为平面向量的复数的高维类似物，他提出了四元数的概念，并且确信四元数对物理学的重要性将不亚于微积分. 他引进了现在常用的微分算子.

$$\nabla = \frac{\partial}{\partial x}\boldsymbol{i} + \frac{\partial}{\partial y}\boldsymbol{j} + \frac{\partial}{\partial z}\boldsymbol{k}.$$

不过这里 $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$ 虽然不是现在我们所用的 $x,y,z$ 轴上的单位向量，却有以下的性质

$\boldsymbol{ij} = \boldsymbol{k}, \boldsymbol{jk} = \boldsymbol{i}, \boldsymbol{ki} = \boldsymbol{j}$ （这一点很像向量积）；

$\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1$ （这是模仿虚数 $\boldsymbol{i}^2 = -1$ 而来的）；

$\boldsymbol{ij} = -\boldsymbol{ji}$ 等等 （这又是模仿向量积）.

于是

$$\nabla(u\boldsymbol{i} + v\boldsymbol{j} + w\boldsymbol{k}) = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) + \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right)\boldsymbol{i}$$

$$+\left(\frac{\partial u}{\partial z}-\frac{\partial w}{\partial x}\right)\boldsymbol{j}+\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)\boldsymbol{k}.$$

第一项是标量部分，除了相差一个符号，就是我们现在说的散度；后三项是向量部分就是我们现在说的旋度.其后伟大的贡献来自麦克斯韦，他在《电磁通论》中系统地使用了向量理论.上述散度、旋度的概念都应归之于他（还有$\nabla^2=-\Delta$，拉普拉斯算子这个名字也是他取的）.其后的发展中起了关键作用有两个人.一是美国的统计物理学家吉卜斯（J.W.Gibbs），他写了第一本未公开发表的《向量分析基础》一书.另一位是英国电机工程师赫维赛德（O.Heaviside）.他的工作完全是为了工程师的需要.因为各方面的需要都很迫切，又缺少统一的名词与记号，多少有一些混乱.在20世纪初数学界的几次国际大会上还专门成立了机构以求统一，然而似乎成果不大.但是有一点很清楚，向量分析，也就是通常的微积分教本中称为"场论"的那一部分，远不只是一种方便的记号，而是真正接触到事物的物理本质.尤其是，它告诉我们，刻画大自然的数学理论决不如人们想像那么"自然"；我们必须讨论不可交换的乘法等等.这不是由于某些人偏爱异想天开的抽象，而是不得不出此途.没有这一切努力，现代的代数学是不可想像的，整个现代数学也是不可想像的.有了这一切，不仅是成就了数学，尤其是成就了物理学.

以上所说与外微分有何关系?例如有一个函数$f(x)$，则$\mathrm{d}f$作为一个向量（以$\{\mathrm{d}x,\mathrm{d}y,\mathrm{d}z\}$为基底）正是 $\mathbf{grad}\ f$.如果$\boldsymbol{A}=(A_x,A_y,A_z)$，同样以$\{\mathrm{d}x,\mathrm{d}y,\mathrm{d}z\}$为基底，则$\mathrm{d}\boldsymbol{A}=\left(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z}\right)\mathrm{d}y\wedge\mathrm{d}z+\left(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x}\right)\mathrm{d}z\wedge\mathrm{d}x+\left(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y}\right)\mathrm{d}x\wedge\mathrm{d}y$，恰好对应于 $\mathbf{curl}\ \boldsymbol{A}$.麦克斯韦指出，$\mathbf{curl\ grad}\ f=0$.于是产生了一个重要问题；若一向量$\boldsymbol{A}$适合$\mathbf{curl}\ \boldsymbol{A}=0$，是否一定存在一个函数$f$使$\boldsymbol{A}=\mathbf{grad}\ f$?如果$\boldsymbol{A}$表示电场，则若$f$存在，$-f$称为这个电场的势，势存在与否是电磁学中重要的问题.麦克斯韦完全理解其重要性，而且知道一般而言势是不存在的.他还指出了这是一个拓扑学问题（麦克斯韦用的是"位置几何学"一语，因为当时还没有"拓扑学"这个术语）.由此可见我们已涉及到科学中很深刻的问题了.建议读者自己去找一下《电磁通论》（武汉出版社有中译本，1992）上册.读一下绪论一章就知道今天我们讲的场论在麦克斯韦时代已发展到什么程度了.

**5．庞加莱引理和上同调理论**　外微分运算$\mathrm{d}$有一个根本之点，即$\mathrm{d}^2\omega=0$对任意阶微分形式$\omega$成立.现在要问，若$\sigma$是微分流形$M$上的一个微分形式，而且$\mathrm{d}\sigma=0$，是否一定有另一微分形式$\omega$存在使$\sigma=\mathrm{d}\omega$?一般说来答案是否定的.§2中曾举了一个例子$x\mathrm{d}y-y\mathrm{d}x$就不是某个函数$\varphi$的微分.这与常微分方程讲的$M\mathrm{d}x+N\mathrm{d}y=0$是否有积分因子的问题有关.在那里我们知道，只要$M$与$N$充分光滑，一定可以找到积分因子$h(x,y)$，使$hM\mathrm{d}x+hN\mathrm{d}y=\mathrm{d}\varphi$，而$\varphi=c$就是这个方程的积分.当然$x\mathrm{d}y-y\mathrm{d}x$并不适合$\mathrm{d}\sigma=0$.但$\sigma=\frac{x\mathrm{d}y-y\mathrm{d}x}{x^2+y^2}$适合$\mathrm{d}\sigma=0$.这时$\sigma=\mathrm{d}\varphi$，而$\varphi(x,y)=\arctan\left(\frac{y}{x}\right)$.但是$\varphi(x,y)$是一个多值函数，而我们在第二章中就已经明确了不应该讨论多值函数.所以仍不能找到一个微分形式$\omega$使$\sigma=\mathrm{d}\omega$.产生这些问题的原因在于出现了一个奇点$(0,0)$.为了系统地处理这些问题，我们先给出

**定义 2**　设$M$为一微分流形，$\omega\in F^p(M)$，$0\leqslant p\leqslant n$，若在$M$上$\mathrm{d}\omega=0$，则称$\omega$为$M$上的（封）闭（closed）形式.若存在$\sigma\in F^{p-1}(M)$使$\omega=\mathrm{d}\sigma$，则称$\omega$为$M$上的恰当（exact）形式.

于是$\mathrm{d}^2=0$即表示凡恰当形式均为闭形式.问题在于其逆是否成立.为了回答这个问题我们

要引入一个几何概念.设 $D\subset\mathbf{R}^n$,$D$ 内有一定点 $P$(不妨设 $P$ 即为原点 $O$).若对任一点 $x\in D$,联结 $P$ 与 $x$ 的线段 $\{tx,0\leqslant t\leqslant 1\}$ 全在 $D$ 内,则 $D$ 称为关于 $P$ 点是星形(starlike)的.这个概念的实质在于区域 $D$ 可以沿着从 0 出发的射线收缩为 $P$ 点,这类可收缩(contractible)区域有许多重要的拓扑学性质.在此情况下我们有

**定理 6**(**庞加莱引理**,Poincaré' s lemma) 设 $M$ 为一微分流形,$M$ 的一个子区域 $S$ 位于 $M$ 之一个坐标邻域 $(U,\varphi)$ 内,且 $\varphi(S)\subset\mathbf{R}^n$ 对某一点为星形的,则任一闭形式 $\omega\in F^p(S)$,$p>0$ 均在 $S$ 内为恰当的.

**证** 不失一般性可以就设 $S\subset\mathbf{R}^n$ 对 $x=0$ 为星形的.$x=(x^1,\cdots,x^n)$ 就是适用于 $S$ 的局部坐标.$\omega\in F^p(S)$,$p>0$.现在我们来找另一个 $(p-1)$ 微分形式 $\sigma$,使在 $S$ 上 $\omega=\mathrm{d}\sigma$.

先看 $\omega$ 为一单项式 $\omega=a(x)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$ 的情况.$\mathrm{d}\omega=0$ 就化为 $\dfrac{\partial a}{\partial x^j}=0$,$j\neq i_1,i_2,\cdots,i_p$.因此在 $S$ 中 $a=a(x_{i_1},\cdots,x_{i_p})$,令 $b$ 为 $\dfrac{\partial b}{\partial x_{i_1}}=a$ 之任一解,容易看到 $\sigma=b\mathrm{d}x^{i_2}\wedge\cdots\wedge\mathrm{d}x^{i_p}$ 就适合所求.这样的解 $\sigma$ 当然不是唯一的.事实上,若 $\tau$ 是任意 $(p-2)$ 微分形式,$\sigma+\mathrm{d}\tau$ 都是这个问题之解.这一个证明可以推广,这样才能看到星形的假设起什么作用.

证明这个定理的关键在于作一个线性算子 $H_p:F^p(M)\to F^{p-1}(M)$,使它具有以下性质:

$$H_{p+1}(\mathrm{d}\omega)+\mathrm{d}H_p(\omega)=\omega. \tag{24}$$

因为若 $\mathrm{d}\omega=0$,则代入此式应得 $\omega=\mathrm{d}H_p(\omega)$.令 $\sigma=H_p(\omega)$ 即得所需解答.$H_p$ 作法如下:

令 $\omega=a_{i_1\cdots i_p}(x)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$,取其任一项,并为简单起见,略去 $a$ 之下标,而不妨设 $\omega=a(x)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$.我们定义算子

图 7-4-1

$$H_p(\omega)=\int_0^1(a(tx)t^{p-1}\mathrm{d}t)\sigma \tag{25}$$

$$\sigma=\sum_{i=1}^p(-1)^{j-1}x^{i_j}\mathrm{d}x^{i_1}\wedge\cdots\wedge\widehat{\mathrm{d}x^{i_j}}\wedge\cdots\wedge\mathrm{d}x^{i_p}. \tag{26}$$

$\mathrm{d}x^{i_j}$ 上的"^"表示这个因子应该删去.星形的假设就用在,当积分变量 $t$ 由 0 变到 1 时,积分域 $\{tx\}$ 就是图 7-4-1 上的线段,而 $a$ 在其上有定义.这样,积分(25)有意义.这个积分对我们并不陌生,因为在求泰勒公式的余项时就遇见过它,那时我们还说过,不必讲到"星形"的概念,这是因为对于 $U\subset\mathbf{R}^n$ 内任一点 $P$(不妨设为原点),恒可以它为心作一个充分小的球 $S$,$S$ 对球心恒为星形的.引入"星形"的名词,是使读者开始与拓扑学中重要的可收缩性概念打交道.同样,算子 $H_p$ 也"大有来头",它是所谓同伦算子的一例.

现在来证(24)式,因为

$$\mathrm{d}\omega=\frac{\partial a(x)}{\partial x^k}\mathrm{d}x^k\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$$

是一个 $(p+1)$ 形式.相应于 $p$ 变成 $p+1$,(26)应成为

$$x^k\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}-\sum_{j=1}^{p}(-1)^{j-1}x^{i_j}\mathrm{d}x^k\wedge\cdots\wedge\widehat{\mathrm{d}x^{i_j}}\wedge\cdots\wedge\mathrm{d}x^{i_p}.$$

这是因为 $x^k$ 起了 $x^{i_0}$ 的作用，而原来 $\mathrm{d}x^{i_j}$ 是第 $j$ 个因子，现在因为多了一个 $\mathrm{d}x^k$ 而成为第 $j+1$ 个因子，所以原来的 $(-1)^{j-1}$ 现在变成了 $(-1)^{(j+1)-1}=-(-1)^{j-1}$. 这样

$$\begin{aligned}H_{p+1}(\mathrm{d}\omega)&=H_{p+1}\left(\frac{\partial a}{\partial x^k}\mathrm{d}x^k\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}\right)\\&=\sum_{h=1}^{n}\left(\int_0^1\frac{\partial a(tx)}{\partial(tx^k)}t^p\mathrm{d}t\right)(x^k\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}-\mathrm{d}x^k\wedge\sigma).\end{aligned}$$

上式的 $\sigma$ 仍是(26)式. 但是

$$\begin{aligned}\mathrm{d}H_p(\omega)&=\mathrm{d}\left[\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\sigma\right]\\&=\mathrm{d}\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\wedge\sigma+\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\mathrm{d}\sigma\\&=\sum_{k=1}^{n}\left(\int_0^1\frac{\partial a(tx)}{\partial(tx^k)}t^p\mathrm{d}t\right)\mathrm{d}x^k\wedge\sigma+\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\mathrm{d}\sigma.\end{aligned}$$

两式相加即有

$$H_{p+1}(\mathrm{d}\omega)+\mathrm{d}H_p(\omega)=\sum_{k=1}^{n}\left(\int_0^1\frac{\partial a(tx)}{\partial(tx^k)}t^p\mathrm{d}t\right)x^k\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}+\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\wedge\mathrm{d}\sigma.$$

但是

$$\sum_{k=1}^{n}x^k\frac{\partial a(tx)}{\partial(tx^k)}=\frac{\partial}{\partial t}a(tx).$$

代入上式第一项并用分部积分法知

$$\begin{aligned}\text{第一项}&=a(tx)t^p\Big|_0^1\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}-p\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}\\&=\omega-p\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}.\end{aligned}$$

第二项中

$$\begin{aligned}\mathrm{d}\sigma&=\sum(-1)^{j-1}\mathrm{d}x^{i_j}\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\widehat{\mathrm{d}x^{i_j}}\wedge\cdots\wedge\mathrm{d}x^{i_p}\\&=\sum\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}=p\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p},\end{aligned}$$

所以

$$\text{第二项}=p\left(\int_0^1a(tx)t^{p-1}\mathrm{d}t\right)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}.$$

两项合并即得定理之证.

庞加莱引理表明，微分流形 $M$ 上的闭形式是否均为恰当形式(从物理上看也就是是否有势存在)，取决于 $M$ 的拓扑性质. 我们在这里所讲的拓扑性质表现为所谓 de Rham 上同调群 $H^*(M)$. 记

$$Z^p(M)=\{M\text{ 上的闭的 }p\text{ 形式}\},$$

并称为 $p$ 维上循环(cocycle)群，这里的“群”是对于 $p$ 形式的加法而言的，因此是阿贝尔(Abel)群(即可换群). 又记

$$B^p(M)=\{M\text{ 上的恰当 }p\text{ 形式}\},$$

并称为 $p$ 维上边缘(coboundary)群.其实 $Z^p(M)$ 和 $B^p(M)$ 都不只是群,而且是线性空间,它们的名称中都有一个"上"字.这也不是偶然的.

显然 $B^p(M)\subset Z^p(M)$,其实这就是 $\mathrm{d}^2=0$.我们也知道了一般说来 $B^p(M)$ 只是 $Z^p(M)$ 的真子空间 —— 真子群.那么至少要问,$Z^p(M)$ 中确非 $B^p(M)$ 的元"实际"上有多少?准确些说,要考虑商空间或商群 $Z^p(M)/B^p(M)$,并且要问,它的维数是多少?有些什么生成元等等.我们就称这个商空间或商群为 $p$ 阶 de Rham 上同调群,并记作 $H^p(M)$:$H^p(M)=Z^p(m)/B^p(M)$.这里 $p\geqslant 0$.$H^p(M)=0$ 就表明 $Z^p(M)=B^p(M)$,亦即每个 $p$ 阶闭形式都是恰当形式.如果 $H^p(M)\neq 0$,它既是线性空间,自当有基底.如果 $H^p(M)$ 是有限维的,而基底为 $\omega_1,\cdots,\omega_r$(当然 $\omega_i$ 都是闭形式),这就是说,任一闭形式 $\omega$ 必可写为 $\omega=\mathrm{d}\sigma+\sum\limits_{i=1}^{r}c_i\omega_i$,如果用 $[\omega]$表示 $\omega$ 在 $H^p(M)$ 中所属的元,则 $[\omega]=\sum\limits_{i=1}^{r}c_i[\omega_i]$.这样,$M$ 上有"多少"个非恰当的闭形式的问题就得到了解决.$H^p(M)$ 的维数是一个拓扑不变量.实际上,若有微分同胚 $\varphi: M\to N$,则 $H^p(M)\cong H^p(N)$,因此其维数自然不变.这个不变量与 $M$ 的其它不变量例如贝蒂数(Betti numbers),欧拉示性数(粗略地说就是 $M$ 有多少个"洞"的个数)都有密切关系.de Rham 上同调理论在物理学(例如量子场论)中有重要的应用,这当然超出了本书的范围,下面来举几个例子.

**例 1** 庞加莱引理告诉我们,若 $M$ 关于其一点是星形的,则 $H^p(M)=0,p>0$,特别是 $\mathbf{R}^n$ 自然是星形域,所以 $H^p(\mathbf{R}^n)=0,p>0$.

$p=0$ 又如何?于此,我们有

**例 2** 若 $M$ 是一个连通流形,可以证明 $H^0(M)\cong\mathbf{R}$.事实上,因为不存在负阶形式,所以 $B^0(M)=0$ 而 $H^0(M)=Z^0(M)$,我们现在来求 $Z^0(M)$.按定义,$Z^0(M)$ 就是闭的 0 阶形式的空间.但 $M$ 上的 0 阶形式就是 $C^\infty(M)$ 函数.$Z^0(M)=\{f\in C^\infty(M);\mathrm{d}f=0\}$.因为 $M$ 是连通的,由 $\mathrm{d}f=0$ 即得 $f=C$ 于$M$ 上,这里 $C\in\mathbf{R}$ 是任意实数,这样

$$H^0(M)=Z^0(M)=\{C\}\cong\mathbf{R}.$$

如果 $M$ 是不连通的,设它有 $r$ 个连通分支,则由 $\mathrm{d}f=0$ 可知 $f$ 在每个连通分支上取常数值.但在不同分支上可以取不同值.这样 $Z^0(M)$ 之每个元都与一组 $r$ 个实数$(C_1,\cdots,C_r)$ 对应,故

$$H^0(M)=Z^0(M)=\{(C_1,\cdots,C_r)\}\cong\mathbf{R}^r.$$

**例 3** 庞加莱引理讲到,可以收缩为一点的流形上,闭形式均为恰当形式,下面看一个明显不能收缩为一点的流形之例:$S^1$.我们前面一再以 $\omega=\dfrac{x\mathrm{d}y-y\mathrm{d}x}{x^2+y^1}$ 为例.它是闭形式但不是恰当的:找不到一个 0 形式(即 $S^1$ 上的函数 $f$) 使 $\omega=\mathrm{d}f$.就是说,向量场 $\left(\dfrac{-y}{x^2+y^2},\dfrac{x}{x^2+y^2}\right)$是没有势函数的 —— 这里讲的是 $S^1$ 上的整体的势函数,而不是 $S^1$ 上任一点的某个邻域中局部的势函数.后者总是存在的,只要这个邻域充分小而且可收缩为一点即可.现在我们用极坐标来考察这个问题.于是 $S^1$ 成为 $r=1,\omega=\mathrm{d}\theta$,(其实 $\mathrm{d}\theta$ 这个记号有毛病,因为 $\theta$ 并非$S^1$ 上的 $C^\infty$ 函数,但这不会影响以下的讨论,因为可以把 $S^1$ 分成几块坐标邻域来处理).$H^0(S^1)\cong\mathbf{R}$ 已如上述.$H^p(S^1)=0$,若 $p>1$,因为 $S^1$ 上没有 $p$ 形式($p>1$,即 $p=2,3,\cdots$),所以 $H^p(S^1)=0,p>$

1. 余下只需讨论 $H^1(S^1)$，$S^1$ 是 1 维流形，所以其切空间的基底是 $\left\{\dfrac{\partial}{\partial\theta}\right\}$，而余切空间的基底是 $\{\mathrm{d}\theta\}$，而 $S^1$ 上所有的闭 1 形式就是 $\omega = \omega(\theta)\mathrm{d}\theta$. 这里 $\omega(\theta)$ 是 $\theta$ 的以 $2\pi$ 为周期的 $C^\infty$ 函数. $\omega$ 为恰当形式 $\mathrm{d}\sigma$，即存在一个 $\theta$ 的以 $2\pi$ 为周期 $C^\infty$ 函数 $\sigma$ 使 $\sigma(\theta) = \int_0^\theta \omega(\theta)\mathrm{d}\theta$，但是上述 $\sigma(\theta)$ 以 $2\pi$ 为周期充分必要条件是

$$\int_0^{2\pi} \omega(\theta)\mathrm{d}\theta = 0.$$

而一般它是不成立的. 为此，我们不考虑 $\omega(\theta)\mathrm{d}\theta$ 而考虑 $[\omega(\theta) - c]\mathrm{d}\theta$，这里 $c$ 应使 $\int_0^{2\pi}[\omega(\theta) - c]\mathrm{d}\theta = 0$ 即 $c = \dfrac{1}{2\pi}\int_0^{2\pi}\omega(\theta)\mathrm{d}\theta$. 因此 $\omega - c\mathrm{d}\theta$ 是恰当形式，即存在 $S^1$ 上的以 $2\pi$ 为周期的 $C^\infty$ 函数 $\sigma(\theta)$ 使

$$\omega - c\mathrm{d}\theta = \mathrm{d}\sigma.$$

这样我们得到了 $S^1$ 上所有的 1 形式的表达式

$$\omega = \mathrm{d}\sigma + c\mathrm{d}\theta.$$

($S^1$ 上所有的 1 形式都是闭的，因为 $S^1$ 是一维的.) 于是 $Z^1(S^1)$ 对于 $B^1(S^1)$ 之等价类 $[\omega] = c\mathrm{d}\theta$;

$$Z^1(\theta)/B^1(S^1) = \{c\mathrm{d}\theta\} \cong \mathbf{R}.$$

于是得出 $H^1(S^1) \cong \mathbf{R}$.

这个例子与磁场可以产生感应电流的奥斯特 – 法拉第的实验有密切的关系，所以我们称 $\omega = \dfrac{-y\mathrm{d}x + x\mathrm{d}y}{x^2 + y^2}$ 为奥斯特 – 法拉第 (Oersted-Faraday) 形式. 我们将在 §6 结束语——麦克斯韦方程组简介中讨论外微分形式与麦克斯韦电磁理论的关系.

**6. 重访古典的向量分析**　上节中我们在引入了霍奇 $*$ 算子后把向量代数中的向量积和混合积用外积表示出来，现在我们将继续利用它给向量分析中几个微分运算公式以更简单的证明.

$*$ 算子的引入与空间中对某一度量下的 o.n. 系的选用有密切关系. 这个 o.n. 系我们记作 $\{e_1, e_2, \cdots, e_n\}$. 因为我们使用了正交坐标系，所以协变与逆变的区别不见了，而我们现在都采用下标. $*$ 算子的公式中经常出现 $p(n-p)+s$，$n$ 是空间维数，在我们的情况下 $n = 3$，$p$ 是微分形式的阶数：$0 \leqslant p \leqslant n$，当 $n = 3$ 时很容易验证，$p(n-p)$ 恒为偶数. 如果我们是在欧氏空间中讨论向量，并选用 $(g_{ij})$ 为正定的度量，从而 $s = 0$，而有 $(-1)^{p(n-p)+s} = 1$. 如果我们在闵可夫斯基四维时空中讨论问题，则 $n = 4$，$s = 3$. 这时 $p(n-p)$ 不一定是偶数，$(-1)^{p(n-p)+s}$ 也不一定是 1.

仿照上一节的记号，凡古典的向量分析中的向量均用箭头或黑体字母表示，而现在则用 o.n. 基 $\{e_i\}$ 表示，如

$$\boldsymbol{A} = A_1\boldsymbol{i} + A_2\boldsymbol{j} + A_3\boldsymbol{k} = A_1e_1 + A_2e_2 + A_3e_3,$$

再回忆一下

$$\begin{aligned}
&*\boldsymbol{i} = *e_1 = e_2 \wedge e_3, \quad *\boldsymbol{j} = *e_2 = e_3 \wedge e_1, \quad *\boldsymbol{k} = *e_3 = e_1 \wedge e_2,\\
&*(e_2 \wedge e_3) = **e_1 = e_1, \quad *(e_3 \wedge e_1) = e_2, \quad *(e_1 \wedge e_2) = e_3, \qquad (27)\\
&*(e_1 \wedge e_2 \wedge e_3) = 1, *1 = e_1 \wedge e_2 \wedge e_3.
\end{aligned}$$

从这里我们还可看到 $**=\mathrm{id}$，所以 $*^{-1}=*$.

现在考虑一个 0 形式 $f$，于是因为 $e_i=\mathrm{d}x_i$

$$\mathrm{d}f=\frac{\partial f}{\partial x_1}\mathrm{d}x_1+\frac{\partial f}{\partial x_2}\mathrm{d}x_2+\frac{\partial f}{\partial x_3}\mathrm{d}x_3=\mathbf{grad}\ f. \tag{28}$$

对于 1 形式 $\boldsymbol{A}=A_1e_1+A_2e_2+A_3e_3=A_1\mathrm{d}x_1+A_2\mathrm{d}x_2+A_3\mathrm{d}x_3$，

$$\mathrm{d}\boldsymbol{A}=\left(\frac{\partial A_3}{\partial x_2}-\frac{\partial A_2}{\partial x_3}\right)\mathrm{d}x^2\wedge\mathrm{d}x^3+\left(\frac{\partial A_1}{\partial x_3}-\frac{\partial A_3}{\partial x_1}\right)\mathrm{d}x^3\wedge\mathrm{d}x^1+\left(\frac{\partial A_2}{\partial x_1}-\frac{\partial A_1}{\partial x_2}\right)\mathrm{d}x^1\wedge\mathrm{d}x^2,$$

所以

$$*\mathrm{d}\boldsymbol{A}=\left(\frac{\partial A_3}{\partial x_2}-\frac{\partial A_2}{\partial x_3}\right)\mathrm{d}x^1+\left(\frac{\partial A_1}{\partial x_3}-\frac{\partial A_3}{\partial x_1}\right)\mathrm{d}x^2+\left(\frac{\partial A_2}{\partial x_1}-\frac{\partial A_1}{\partial x_2}\right)\mathrm{d}x^3=\mathbf{curl}\ \boldsymbol{A}. \tag{29}$$

又因

$$*\boldsymbol{A}=A_1\mathrm{d}x^2\wedge\mathrm{d}x^3+A_2\mathrm{d}x^3\wedge\mathrm{d}x^1+A_3\mathrm{d}x^1\wedge\mathrm{d}x^2,$$

所以

$$*\mathrm{d}*\boldsymbol{A}=*\left[\left(\frac{\partial A_1}{\partial x_1}+\frac{\partial A_2}{\partial x_2}+\frac{\partial A_3}{\partial x_3}\right)\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3\right]=\frac{\partial A_1}{\partial x_1}+\frac{\partial A_2}{\partial x_2}+\frac{\partial A_3}{\partial x_3}=\mathrm{div}\ \boldsymbol{A}. \tag{30}$$

于是，梯度，旋度与散度都可以用外微分来表示. 由此，外微分的性质将直接给出这些量的相关性质. 例如从定理 3 的性质 (i)，即 (18) 式很容易地导出了以下诸多很容易混淆的公式

(i) $\mathrm{div}(\mathbf{curl}\ \boldsymbol{A})=0$，实即 (31)

$$*\mathrm{d}*(*\mathrm{d}\boldsymbol{A})=*\mathrm{d}^2\boldsymbol{A}=0\quad(**\omega=\omega).$$

(ii) $\mathbf{curl}(\mathbf{grad}\ f)=0$，即是 (32)

$$*\mathrm{d}(\mathrm{d}f)=*\mathrm{d}^2f=0.$$

(iii) $\mathbf{curl}(f\boldsymbol{A})=f\mathbf{curl}\ \boldsymbol{A}+\mathbf{grad}\ f\times\boldsymbol{A}$，即是 (33)

$$\begin{aligned}*\mathrm{d}(f\boldsymbol{A})&=f*\mathrm{d}\boldsymbol{A}+*(\mathrm{d}f\wedge\boldsymbol{A}).\\&=f\mathbf{curl}\ \boldsymbol{A}+\mathbf{grad}\ f\times\boldsymbol{A}.\end{aligned}$$

(iv) $\mathrm{div}(f\boldsymbol{A})=f\mathrm{div}\ \boldsymbol{A}+\mathbf{grad}\ f\cdot\boldsymbol{A}$，因为 (34)

$$\begin{aligned}*\mathrm{d}(*f\boldsymbol{A})&=f(*\mathrm{d}*\boldsymbol{A})+*(\mathrm{d}f\wedge*\boldsymbol{A})\\&=f\mathrm{div}\ \boldsymbol{A}+\mathbf{grad}\ f\times\boldsymbol{A}.\end{aligned}$$

(v) $\mathrm{div}(\boldsymbol{A}\times\boldsymbol{B})=\boldsymbol{B}\cdot\mathbf{curl}\ \boldsymbol{A}-\boldsymbol{A}\cdot\mathbf{curl}\ \boldsymbol{B}$，因为 (35)

$$\begin{aligned}*\mathrm{d}*(\boldsymbol{A}\times\boldsymbol{B})&=*\mathrm{d}*(*\boldsymbol{A}\wedge\boldsymbol{B})=*\mathrm{d}(\boldsymbol{A}\wedge\boldsymbol{B})\\&=*(\mathrm{d}\boldsymbol{A}\wedge\boldsymbol{B}-\boldsymbol{A}\wedge\mathrm{d}\boldsymbol{B})\\&=*(\mathbf{curl}\ \boldsymbol{A}\wedge\boldsymbol{B}-\boldsymbol{A}\wedge\mathbf{curl}\ \boldsymbol{B})=\boldsymbol{B}\cdot\mathbf{curl}\ \boldsymbol{A}-\boldsymbol{A}\cdot\mathbf{curl}\ \boldsymbol{B}.\end{aligned}$$

我们特别注意的当然还是庞加莱引理的应用. 为此，假定都是在某一星形区域或一点的邻域中考虑问题.

首先设 $\mathbf{curl}\ \boldsymbol{A}=0$，亦即 $*\mathrm{d}\boldsymbol{A}=0$ 而有 $\mathrm{d}\boldsymbol{A}=0$. 于是一定存在一个 0 形式即 $C^\infty$ 函数 $f$ 使

$$\boldsymbol{A}=\mathbf{grad}\ f.$$

在物理上，一个向量场 $\boldsymbol{A}$ 若适合 $\mathbf{curl}\ \boldsymbol{A}=0$ 的条件，就称为无旋场. 于是上面的结论就是无旋场一定有势. 但是这只限于区域为星形或为某一点之邻域时成立，否则势将会是多值函数. 例 3 中

的奥斯特 – 法拉第形式（或称奥斯特 – 法拉第场）就是这样，因为这时有

$$\omega = \mathrm{d}(\sigma - c\theta),$$

而 $\theta$ 化到直角坐标系以后成为 $\arctan\dfrac{y}{x}$ 就是一个多值函数.

另外，如果 $\operatorname{div}\boldsymbol{A} = 0$，亦即 $*\mathrm{d}(*\boldsymbol{A}) = 0$，因此 $*\boldsymbol{A}$ 是一个闭 2 形式，这时一定存在一个 1 形式 $\boldsymbol{B}$ 使 $*\boldsymbol{A} = \mathrm{d}\boldsymbol{B}$，或 $\boldsymbol{A} = *\mathrm{d}\boldsymbol{B}$. 用古典的向量分析语言来表达就是：若 $\operatorname{div}\boldsymbol{A} = 0$，则必存在一个向量 $\boldsymbol{B}$（称为向量势）使 $\boldsymbol{A} = \mathbf{curl}\,\boldsymbol{B}$.

讲到向量分析就不能不讲一下拉普拉斯算子，读者们可能都熟悉 $\Delta u = \operatorname{div}\mathbf{grad}\,u$，而 $\Delta u = 0$ 是古典的数学物理中可说是最重要的算子. 但是应该指出，它在微分流形上有含意深远的推广. 不只是因为它不再是常系数偏微分算子，而且不只是作用在 0 形式（即函数）上，而可以作用在一般的 $p$ 形式上. 我们也习惯把它称为 Laplace-Beltrami 算子. 它的理论与霍奇的理论和 de Rham 上同调理论密切相关，而且是现代的数学物理的基础之一部分. 作用在 0 形式上的拉普拉斯算子以外微分算子 d 为基础，而且要把 d 的“对偶”也同时考虑进来.

我们知道 $\mathrm{d}: F^p(M) \to F^{p+1}(M)$，它的“对偶”$\delta$：则会使微分形式的次数下降一次：$\delta$ 由下面的可变换图式来定义：

$$\varepsilon = (-1)^p,$$

$$\begin{array}{ccc} F^p(M) & \xrightarrow{\varepsilon\delta} & F^{p-1}(M) \\ {\scriptstyle *}\downarrow & & \downarrow{\scriptstyle *} \\ F^{n-p}(M) & \xrightarrow[\mathrm{d}]{} & F^{n-p+1}(M) \end{array} \tag{36}$$

所谓可交换图式即是如果有两条路径起点与终点相同，则沿每一条道径的算子之积（靠近起点的算子先作用），结果相同. 因为 (36) 中有两条路径由左上方的 $F^p(M)$ 通向右上方的 $F^{p-1}$，其一是沿水平方向的 $\varepsilon\delta$，另一条是先用 $*$ 向下通向 $F^{n-p}(M)$，再用 d 向右走到 $F^{n-p+1}(M)$，最后再用霍奇算子的逆 $*^{-1}$ 向上通到 $F^{p-1}(M)$. 所以

$$\delta = (-1)^p *^{-1} \circ \mathrm{d} \circ *,$$
$$\delta: F^p(M) \to F^{p-1}(M).$$

以上我们令 $p \geqslant 1$，否则 $F^{p-1}(M)$ 无法定义. 除上式外，我们再补充定义 $\delta\omega = 0, \omega \in F^0(M)$，于是最后得到，对于 $\omega \in F^p(M)$，

$$\delta\omega = \begin{cases} (-1)^p *^{-1} \circ \mathrm{d} \circ *, & \varepsilon = (-1)^p, p > 0, \\ 0 & p = 0. \end{cases} \tag{37}$$

$\delta$ 是 d 的对偶算子. 为什么是对偶，我们在讲了积分理论以后就清楚了.

现在我们定义 $\mathbf{R}^n$ 上的 Laplace-Beltrami 算子为

$$\Delta\omega = (\mathrm{d}\delta + \delta\mathrm{d})\omega, \quad \omega \in F^p(\mathbf{R}^n), p \geqslant 0. \tag{38}$$

古典的拉普拉斯算子是 $F^0(\mathbf{R}^n)$（即 $C^\infty$ 函数）空间上的算子（当然可以推广到甚至是广义函数空间上去）而现在则是由微分形式空间 $F^p(\mathbf{R}^n)$ 到 $F^p(\mathbf{R}^n)$ 上的映射. 当 $p = 0$ 时，注意到这时 $\delta\omega = 0, \omega \in F^0(\mathbf{R}^n)$，

$$\Delta\omega = \delta\mathrm{d}\omega = [-*^{-1} \circ \mathrm{d} \circ * \circ \mathrm{d}]\omega = -[* \circ \mathrm{d} \circ * \circ \mathrm{d}]\omega.$$

这里我们利用了上节(39)式：在 $\mathbf{R}^n$ 中对于欧几里得度量 $p=0$ 时 $\delta=0$，故 $**=\mathrm{id}$，于是

$$\Delta\omega=-\operatorname{div}\mathbf{grad}\,\omega.$$

可是现在的 Laplace-Beltrami 算子与古典的拉普拉斯算子相差一个符号，这使它应用起来更方便. 因此在现代文献中我们常见到

$$\Delta=-\left[\left(\frac{\partial}{\partial x^1}\right)^2+\cdots+\left(\frac{\partial}{\partial x^n}\right)^2\right].$$

我们还要问，对 $p>0$ 的 $F^p(\mathbf{R}^3)$，$\Delta\omega$ 是什么？一个常用的情况是 $\omega\in F^1(\mathbf{R}^3)$，我们就记为 $\omega=\boldsymbol{u}$. 这时，因 $\mathrm{d}\omega\in F^2(\mathbf{R}^3)$，故

$$\delta\mathrm{d}\omega=*\mathrm{d}*\mathrm{d}\omega=\mathbf{curl}\ \mathbf{curl}\ \boldsymbol{u}.$$

而

$$\mathrm{d}\delta\omega=-\mathrm{d}*\mathrm{d}*\omega=-\mathbf{grad}\operatorname{div}\boldsymbol{u}.$$

另一方面我们可以直接计算 $\Delta\omega$ 如下：不妨设 $\omega=u_i\mathrm{d}x^i$，于是经过不太复杂的计算有 $\Delta\omega=-\left[\left(\frac{\partial}{\partial x^1}\right)^2+\left(\frac{\partial}{\partial x^2}\right)^2+\left(\frac{\partial}{\partial x^3}\right)^2\right]\boldsymbol{u}$，即对 $\boldsymbol{u}$ 之每个分量作古典的拉普拉斯算子并反号. 于是有

$$\mathbf{curl}\ \mathbf{curl}\ \boldsymbol{u}=\mathbf{grad}\operatorname{div}\boldsymbol{u}-\nabla^2\boldsymbol{u}.\tag{39}$$

这里我们采用了古典的微积分教本中常用的记法：$\nabla^2=\nabla\cdot\nabla$，而$\nabla=\frac{\partial}{\partial x^1}\boldsymbol{i}+\frac{\partial}{\partial x^2}\boldsymbol{j}+\frac{\partial}{\partial x^3}\boldsymbol{k}$，这样记就不会对 $\Delta$ 的符号产生误会了. (39) 式如果按古典的讲法来证明，并不是一件容易事.

上面我们利用外微分形式特别是霍奇 $*$ 算子，非常清楚地讲了古典的"场论"在 $\mathbf{R}^n$ 中的许多公式. 但是如果换其它坐标系，更容易看出它的好处来. 上一节已经看得很明白，在任一度量 $(g_{ij})$ 下，只要找到了 o.n. 系 $\{e^1,\cdots,e^n\}$ 就可以定义 $*$ 算子，因此也就可能定义 **grad**，**curl** 与 div 等等. 在 o.n. 基底下，度量矩阵 $(g_{ij})$ 一定是对角矩阵，这是因为

$$g_{ij}=\langle e_i,e_j\rangle,$$

所以 $i\neq j$ 时 $g_{ij}=0$. 这样

$$\mathrm{d}s^2=H_1(\mathrm{d}x^1)^2+\cdots+H_n(\mathrm{d}x^n)^2.\tag{40}$$

如果取度量为正定的，则有 $H_i>0$，而且 $H_1\cdots H_n=g=\det(g_{ij})>0$. 所以只要令

$$e^i=\sqrt{H_i}\,\mathrm{d}x^i,\quad e_i=\frac{1}{\sqrt{H_i}}\frac{\partial}{\partial x^i},$$

就可以得到所需的 o.n. 系，而可对它们定义 $*$ 算子. 下面我们限于三维的具有正定度量的流形 $M$，即三维黎曼流形，并以 $(x^1,x^2,x^3)$ 为其局部坐标. 所以有时我们也把 $M$ 直接写成 $\mathbf{R}^3$.

于是对于 $\omega\in F^0(M)$（我们习惯用 $f$ 表示一个 0 微分形式，因为它确实就是一个函数），

$$\mathrm{d}f=\frac{\partial f}{\partial x^i}\mathrm{d}x^i=\frac{1}{\sqrt{H_i}}\frac{\partial f}{\partial x^i}e^i,$$

因此我们应定义

$$\mathbf{grad}\ f=\left(\frac{1}{\sqrt{H_1}}\frac{\partial f}{\partial x^1},\cdots,\frac{1}{\sqrt{H_n}}\frac{\partial f}{\partial x^n}\right).\tag{41}$$

往下，如果 $\omega\in F^1(M)$，$\omega=A_ie^i$ 或者写为 $\omega=\boldsymbol{A}$，我们有

$$\omega=(A_i\sqrt{H_i})\mathrm{d}x^i.$$

在现在的情况有

$$\begin{aligned}\mathrm{d}\omega = &\left(\frac{\partial A_3\sqrt{H_3}}{\partial x^2}-\frac{\partial A_2\sqrt{H_2}}{\partial x^3}\right)\mathrm{d}x^2\wedge\mathrm{d}x^3\\ &+\left(\frac{\partial A_1\sqrt{H_1}}{\partial x^3}-\frac{\partial A_3\sqrt{H_3}}{\partial x^1}\right)\mathrm{d}x^3\wedge\mathrm{d}x^1\\ &+\left(\frac{\partial A_2\sqrt{H_2}}{\partial x^1}-\frac{\partial A_1\sqrt{H_1}}{\partial x^2}\right)\mathrm{d}x^1\wedge\mathrm{d}x^2.\end{aligned}$$

而

$$\mathrm{d}\omega = \frac{1}{\sqrt{H_2H_3}}\left(\frac{\partial A_3\sqrt{H_3}}{\partial x^2}-\frac{\partial A_2\sqrt{H_2}}{\partial x^3}\right)e^2\wedge e^3+\cdots. \tag{42}$$

因此，我们应该有

$$\mathbf{curl}\ \boldsymbol{A} = \frac{1}{\sqrt{H_2H_3}}\left(\frac{\partial A_3\sqrt{H_3}}{\partial x^2}-\frac{\partial A_2\sqrt{H_2}}{\partial x^3}\right)e^1+\cdots. \tag{43}$$

下面没有写出的项均可由第一项用轮换对称得出. 最后看 $\mathrm{div}\ \boldsymbol{A} = *\mathrm{d}*\boldsymbol{A}$，因为

$$\begin{aligned}*\boldsymbol{A} &= A_1e^2\wedge e^3+A_2e^3\wedge e^1+A_3e^1\wedge e^2\\ &= A_1\sqrt{H_2H_3}\mathrm{d}x^2\wedge\mathrm{d}x^3+A_2\sqrt{H_3H_1}\mathrm{d}x^3\wedge\mathrm{d}x^1+A_3\sqrt{H_1H_2}\mathrm{d}x^1\wedge\mathrm{d}x^2,\\ \mathrm{d}*\boldsymbol{A} &= \left(\frac{\partial}{\partial x_1}A_1\sqrt{H_2H_3}+\frac{\partial}{\partial x^2}A_2\sqrt{H_3H_1}+\frac{\partial}{\partial x^3}A_3\sqrt{H_1H_2}\right)\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3\\ &= \frac{1}{\sqrt{H_1H_2H_3}}\left(\frac{\partial}{\partial x^1}A_1\sqrt{H_2H_3}+\cdots\right)e^1\wedge e^2\wedge e^3.\end{aligned}$$

所以

$$\mathrm{div}\ \boldsymbol{A} = \frac{1}{\sqrt{H_1H_2H_3}}\left(\frac{\partial}{\partial x^1}A_1\sqrt{H_2H_3}+\cdots\right). \tag{44}$$

最后我们来看 Laplace-Beltrami 算子，对于 $f\in F^0(M)$，有

$$\begin{aligned}\Delta f &= (\mathrm{d}\delta+\delta\mathrm{d})f = \delta\mathrm{d}f = -*\mathrm{d}*\mathrm{d}f\\ &= -\mathrm{div}\ \mathbf{grad}\ f\\ &= -\frac{1}{\sqrt{H_1H_2H_3}}\left[\frac{\partial}{\partial x^1}\sqrt{\frac{H_2H_3}{H_1}}\frac{\partial f}{\partial x^1}+\cdots\right].\end{aligned} \tag{45}$$

这些公式有一个常见的用处，即在 $\mathbf{R}^3$ 中引入球坐标

$$x^1 = r\sin\theta\cos\varphi,\ x^2 = r\sin\theta\sin\varphi,\ x^3 = r\cos\theta$$

时，很容易算出

$$\mathrm{d}s^2 = \mathrm{d}r^2+r^2\mathrm{d}\theta^2+r^2\sin^2\theta\mathrm{d}\varphi^2.$$

不妨以 $(r,\theta,\varphi)$ 为 $(x^1,x^2,x^3)$，现在我们知道不应该令

$$\mathbf{grad}\ f = \left(\frac{\partial f}{\partial r},\frac{\partial f}{\partial\theta},\frac{\partial f}{\partial\varphi}\right),$$

因为这个“记号”没有上面这些公式表示的几何性质，而应该注意现在应可得

$$H_1 = 1,\ H_2 = r^2,\ H_3 = r^2\sin^2\theta,$$

而有

$$\mathbf{grad}\ f = \left(\frac{\partial f}{\partial r}, \frac{1}{r}\frac{\partial f}{\partial \theta}, \frac{1}{r\sin\theta}\frac{\partial f}{\partial \varphi}\right). \tag{46}$$

对于一个向量 $\boldsymbol{A}$，在坐标系 $(r, \theta, \varphi)$ 中应该先把它写成

$$\boldsymbol{A} = A_1 \mathrm{d}r + A_2 (r\mathrm{d}\theta) + A_3 (r\sin\theta)\mathrm{d}\varphi,$$

然后用(43)和(44)可得

$$\begin{aligned}\mathbf{curl}\ \boldsymbol{A} = &\frac{1}{r^2\sin\theta}\left(\frac{\partial A_3(r\sin\theta)}{\partial\theta} - \frac{\partial A_2 r}{\partial\varphi}\right)\mathrm{d}r + \\ &\frac{1}{\sin\theta}\left(\frac{\partial A_1}{\partial\varphi} - \frac{\partial A_3(r\sin\theta)}{\partial r}\right)\mathrm{d}\theta + \\ &\sin\theta\left(\frac{\partial A_2 r}{\partial r} - \frac{\partial A_1}{\partial\theta}\right)\mathrm{d}\varphi.\end{aligned} \tag{47}$$

$$\mathrm{div}\ \boldsymbol{A} = \frac{1}{r^2\sin\theta}\left(\frac{\partial}{\partial r}A_1 r^2\sin\theta + \frac{\partial}{\partial\theta}A_2 r\sin\theta + \frac{\partial}{\partial\varphi}A_3 r\right). \tag{48}$$

而拉普拉斯算子(注意符号)是

$$\begin{aligned}\Delta f &= -\frac{1}{r^2\sin\theta}\left[\left(\frac{\partial}{\partial r}r^2\sin\theta\frac{\partial f}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{\partial}{\partial\varphi}\left(\frac{1}{\sin\theta}\frac{\partial f}{\partial\varphi}\right)\right] \\ &= -\frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\varphi^2}\right].\end{aligned} \tag{49}$$

# §5　微分形式在流形上的积分

**1. 微分形式在 $\mathbf{R}^n$ 上的积分**　本节中我们将讨论一个微分流形 $M$ 上的微分形式 $\omega \in F^n(M)$（$\omega$ 的阶数与 $M$ 的维数相同）在 $M$ 上的积分问题. 因为 $M$ 是光滑的，$F^n(M)$ 中的元素为 $\wedge^n(M)$ 的 $C^\infty$ 切口，因此也是光滑的. 在光滑的积分区域上对光滑的被积表达式作积分，黎曼积分就已经够用了. 所以本章中的积分都限于黎曼积分而不去讨论第四章中的勒贝格积分，也就不必讨论 $M$ 上如何定义测度的问题. 当然这不等于说在研究有关流形的几何问题时用不着勒贝格积分. 退一步说，至少要有黎曼度量才行. 一个微分流形，如我们所已假设的，是豪斯多夫空间，而且满足第二可数性公理，在这个条件下，这个流形上一定有黎曼度量存在(证明见定理 6)，也可以证明它由此就是一个度量空间(证明略去). 但是我们现在还没有利用这些结果. 在这种情况下如何定义 $M$ 上的积分?办法就是利用 $M$ 上的局部坐标，把积分区域映射到 $\mathbf{R}^n$ 上去. 这个映射尽管"我们说它是可微的"，却不涉及测度问题. 设 $U \subset M$ 是一个坐标邻域，$\varphi: U \to \mathbf{R}^n$ 是坐标映射，记 $\varphi(U) = V$，$\omega$ 在 $\varphi^*$ 下之像为 $\Omega$，则

$$\Omega = a(x)\mathrm{d}x^l \wedge \cdots \wedge \mathrm{d}x^n.$$

可是 $\omega$ 是 $\Omega$ 在 $\varphi$ 下的拉回：$\omega = \varphi^*\Omega$，于是我们定义

$$\int_U \omega = \int_U \varphi^*\Omega = \int_{\varphi(U)} \Omega = \int_V a(x)\mathrm{d}x^1\cdots\mathrm{d}x^n, \tag{1}$$

这就完全绕过了测度问题.

但是若在 $U$ 中有了另一个局部坐标 $\psi$: $(y^1,\cdots,y^n)$，不失一般性，不妨设 $y$ 与 $x$ 坐标有相同定向. 如果在新坐标系中 $\omega$ 成为

$$\tilde{\Omega}=b(y)\mathrm{d}y^1\wedge\cdots\wedge\mathrm{d}y^n,$$

则按 (1) 式，$\int_U\omega$ 应定义为

$$\int_U\omega=\int_\sigma\psi^*\ \tilde{\Omega}=\int_{\psi(U)}\tilde{\Omega}=\int_V b(y)\mathrm{d}y^1\cdots\mathrm{d}y^n. \tag{2}$$

但由外微分形式的性质

$$b(y)\mathrm{d}y^1\wedge\cdots\wedge\mathrm{d}y^n=b(y(x))\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n$$
$$=a(x)\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n.$$

即有 $a(x)=b(y(x))J$，$J=\dfrac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}>0$，因为 $y$ 与 $x$ 有相同定向，所以 $J=|J|$，而由黎曼积分中重积分之变量变换公式

$$\int_V a(x)\mathrm{d}x^1\cdots\mathrm{d}x^n=\int_V b(y(x))|J|\mathrm{d}x^1\cdots\mathrm{d}x^n$$
$$=\int_V b(y)\mathrm{d}y^1\cdots\mathrm{d}y^n.$$

因此这个定义是与坐标选择无关的，因而是合理的. 研究微分流形上的积分理论时一个关键之点是要使所得的结果与坐标的选择无关，微分形式的性质正好保证了这一点.

尽管我们给出了 (1) 或 (2) 作为 $\omega$ 在 $U\subset M$ 上积分的定义的基础，其中仍有值得推敲的地方. 因为例如对于 (1) 式，要求 $\int_V a(x)\mathrm{d}x^1\cdots\mathrm{d}x^n$ 有意义，应该要求 $\partial V$ 是勒贝格零测度集（第四章 §1)，而且对 $a(x)$ 之光滑性不能只限于要求它在开集 $V$ 中光滑，因为这时 $a(x)$ 在 $\partial V$ 上仍可能很不规则，使得 $a(x)$ 在 $V$ 上不一定可积（即所谓反常积分）. 再说，在许多问题中 $\partial V$ 不一定是光滑的，例如它可以有边、有角等等，而处理这类问题总是很麻烦的. 因此，我们干脆回到 $\mathbf{R}^n$，在其上重新建立起微分形式的积分理论，然后再来讨论如何把这些结果搬到微分流形上来.

在讲黎曼积分时我们常将积分区域分割成小的 $n$ 维长方体（以下为简单起见，就称它们为矩形，正如在第四章中称它们为区间一样）. 现在我们要采取一种更有系统的分割方法，即分割一个空间为"单形"(simplex). 这不但是为了使读者有机会和拓扑学中的某些基本概念和方法早打交道，更是为了使读者注意到积分理论的组合学的侧面，这一点我们在第四章中已初步强调过了.

为了介绍什么是单形，先介绍一下线性空间 $V$ 中的仿射子空间 $A$，如果有 $V$ 的一个线性子空间 $W\subset V$，使

$$A=x_0+W,$$

则称 $A$ 为 $V$ 之仿射子空间. 从几何上看，仿射子空间无非是把一个线性子空间 $W$ 的原点移到 $x_0$ 点，这在几何上和物理上都是自然的. 例如，以地球或以太阳为宇宙之中心均无不可. 但仿射子空间在运算上显然有些麻烦：如果 $y_1,y_2\in A$，则 $y_1-y_2\in W$ 而不一定在 $A$ 中，$y_1+y_2$ 更不一定能写为 $x_0+$ ($W$ 之某个元)，除非 $x_0\in W$，而这时也就无所谓仿射空间 $A$ 了. 今取 $W$ 的一个基底

$\{W_1,\cdots,W_k\}$，并令 $x_i = x_0 + W_i$. 当 $\{W_1,\cdots,W_k\}$ 线性无关时，由 $\sum_{i=1}^{k}\lambda_i W_i = 0$，亦即由 $\sum_{i=1}^{k}\lambda_i(x_i - x_0) = \sum_{i=0}^{k}\lambda_i x_i = 0$ 时可得 $\lambda_1 = \cdots = \lambda_k = 0$，这里 $\lambda_0 = -(\lambda_1 + \cdots + \lambda_k)$，从而 $\sum_{i=0}^{k}\lambda_i = 0$，因此由 $\{W_1,\cdots,W_n\}$ 之线性无关（即在线性空间意义下无关），关于 $\{x_0,x_1,\cdots,x_k\}$ 可得以下性质：若 $\sum_{i=0}^{k}\lambda_i x_i = 0$ 以及 $\sum_{i=0}^{k}\lambda_i = 0$ 可得 $\lambda_0 = \cdots = \lambda_k = 0$. 这个性质我们说成是 $\{x_0,x_1,\cdots,x_k\}$ 仿射无关（即在仿射空间意义下无关）. 反之由 $\{x_0,x_1,\cdots,x_k\}$ 仿射无关，必可得 $\{x_0 - x_{i_0}, x_1 - x_{i_0},\cdots,\widehat{x_{i_0} - x_{i_0}},\cdots,x_k - x_{i_0}\}$ 线性无关. $\{x_0,x_1,\cdots,x_k\}$ 称为 $A$ 的一个仿射基底，因而仿射空间 $A$ 中的元一定可以表示为某个仿射基底的仿射组合如下：任取 $x\in A$，必有 $W$ 中一个元（它可以表示为 $\sum_{i=1}^{k}\mu_i W_i = \sum_{i=1}^{k}\mu_i(x_i - x_0)$），使

$$x = x_0 + \sum_{i=1}^{k}\mu_i(x_i - x_0) = \mu_0 x_0 + \sum_{i=1}^{k}\mu_i x_i$$
$$= \sum_{i=0}^{k}\mu_i x_i, \qquad \mu_0 = 1 - \sum_{i=1}^{k}\mu_i. \tag{3}$$

形如(3)的表达式称为 $\{x_0,x_1,\cdots,x_k\}$ 的仿射组合，但

$$\sum_{i=0}^{k}\mu_i = 1 \qquad (x_0,x_1\in V). \tag{4}$$

最简单的例子是 $\{x_0,x_1\}$ 的仿射组合之集，即

$$L = \{\mu x_0 + (1-\mu)x_1;\ \mu\in\mathbf{R}\}.$$

它就是过 $x_0,x_1$ 两点的直线，它与 $V$ 的线性子空间不同，它不必经过原点. 如果在 $L$ 中限制 $0\leqslant\mu\leqslant 1$，则将得到 $L$ 上由 $x_0$ 到 $x_1$ 的线段（端点包括在内）.

有了仿射组合的概念以后，我们就可以定义单形了.

**定义1** $n$ 维线性空间 $V$ 中的 $k$ 维单形（这里 $0\leqslant k\leqslant n$），简称为 $k$ 单形（k-simplex），即 $k+1$ 个仿射无关的点 $\{x_0,x_1,\cdots,x_k\}$ 之如下的仿射组合的集合.

$$\sigma_k = \langle x_0,x_1,\cdots,x_k\rangle = \left\{x;\ x = \sum_{i=0}^{k}\lambda_i x_i, 0\leqslant\lambda_i\leqslant 1, \sum_{i=0}^{k}\lambda_i = 1\right\}. \tag{5}$$

这些点称为 $\sigma_k$ 的顶点.

以上受了限制 $0\leqslant\lambda_i\leqslant 1$ 的仿射组合是一个凸集合. 因此 $\sigma_k$ 就是由这些顶点所生成的闭凸集——或称凸包. $\lambda_i$ 称为 $x$ 点的第 $i$ 个重心坐标（barycentric coordinates），我们常记之为 $b_i(x) = \lambda_i$. 注意，$\sigma_k\subset V$ 乃是 $k$ 维空间的一个子集，所以其中的点的独立的坐标应有 $k$ 个分量，而现在有 $k+1$ 个 $\lambda_i$，所以它们并非本来意义的坐标，好在它们也不是独立的，因为有 $\sum_{i=0}^{k}\lambda_i = 1$.

0 单形就是一个点 $\{x_0\}$；1 单形是一个线段；2 单形是一个三角形. 在考虑积分问题时必须考虑积分区域的定向，所以我们要规定单形(5)的定向. 仿照 $\mathbf{R}^n$ 中定向即坐标按一定次序的排列，现在 $\sigma_k$ 的定向也就是把顶点 $\{x_0,x_1,\cdots,x_k\}$ 按一定次序的排列. 以下 $\langle x_0,x_1,\cdots,x_h\rangle$ 恒表示顶点已按此顺序排列的有向单形. 如果将顶点按另一顺序排列，则有一个 $\{0,1,\cdots,k\}$ 的排列 $\pi$，而

排列后的单形成为 $\langle x_{\pi(0)}, x_{\pi(1)}, \cdots, x_{\pi(k)}\rangle$. 如果 $\pi$ 是一个偶(奇)排列，就说 $\langle x_{\pi(0)}, x_{\pi(1)}, \cdots, x_{\pi(k)}\rangle$ 与 $\langle x_0, x_1, \cdots, x_k\rangle$ 有相同定向(相反定向，并记 $\langle x_{\pi(0)}, x_{\pi(1)}, \cdots, x_{\pi(k)}\rangle = -\langle x_0, x_1, \cdots, x_k\rangle$). 总之

$$\langle x_{\pi(0)}, x_{\pi(1)}, \cdots, x_{\pi(k)}\rangle = \operatorname{sgn}\pi\langle x_0, x_1, \cdots, x_k\rangle.$$

每一个单形都有边缘. 设有单形 $\sigma = \langle x_0, x_1, \cdots, x_k\rangle$，则从这些顶点中任取若干个组成一个新单形(其维数可从 0—— 即由单个顶点 $\{x_i\}$ 构成)一直到 $k$(就是 $\sigma_k$ 自身). 设某一个这样的新单形记作 $T_l$($l$ 表示其维数)，则称 $T_l$ 为 $\sigma_k$ 的 $l$ 维面(特别是 $l < k$ 时称为真面). 如果 $\tau$ 是 $\sigma$ 的一个面，我们记为 $\tau \prec \sigma$. 我们特别要注意 $\sigma_k$ 的低一维的 $(k-1)$ 维面 $\sigma_{k-1}^{(i)} = \langle x_0, x_1, \cdots, \hat{x}_i, \cdots, x_k\rangle$，$\hat{x}_i$ 表示删去 $x_i$. 更低维的面因为都是 $\sigma_{k-1}^{(i)}$ 的面，而单形都是闭的，所以这些面都已含于 $\sigma_{k-1}^{(i)}$ 中了. 我们本可就用上述的顶点次序安排 $\sigma_{k-1}^{(i)}$ 之定向. 但由于一些几何上直观可见的理由，我们用 $(-1)^i\sigma_{k-1}^{(i)}$ 作为其定向. 这样一来，我们就可以定义单形 $\sigma_k$ 的边缘算子 $\partial_k$(或简记为 $\partial$) 如下：

$$\partial\sigma_k = \sum_{i=0}^{k}(-1)^i\langle x_0, \cdots, \hat{x}_i, \cdots, x_k\rangle. \tag{6}$$

我们想要研究的几何图形是由有限个单形 $\{\sigma^1, \cdots, \sigma^m\}$ 组成的，这种几何图形叫做复形(准确些称为单纯复形(simplicial complex)) $K$，对复形有以下要求：

1. $K$ 是由有限个单形组成的，每个单形的维数不一定相同.

2. 若 $\sigma$ 是 $K$ 中一个单形，则 $\sigma$ 的所有的面也在 $K$ 中.

3. 若 $\tau$ 与 $\sigma$ 均为 $K$ 中的单形，则或者 $\tau \cap \sigma = \varnothing$，或者 $\tau \cap \sigma$ 是 $\tau$ 与 $\sigma$ 的公共面. 这种情况下我们说 $\tau$ 与 $\sigma$ 规则相处. 图 7－5－1 是一些不规则相处的单形之例.

我们求积分的区域可以是复形中若干个单形组成的，但每个单形出现的次数不一定是一次，而且定向也可改变. 因此我们将考虑一个"链"(chain)，即

$$c = \sum n_i\sigma^i, \qquad n_i \text{ 为整数}. \tag{7}$$

而因为一个复形中只有有限个单形，所以只有有限多个 $n_i \neq 0$，而且我们规定 $\sigma^i$ 之维数相同，这个公共的维数也就成为链的维数. (7) 式是一个形式和，按照形式的加法，它将构成一个可交换群. 我们要求当且仅当 (7) 式中一切 $n_i = 0$ 时，这个链才是上述群中的零元. 若上述公共维数为 $k$，则称此群是一个 $k$ 维链群，记作 $C_k$.

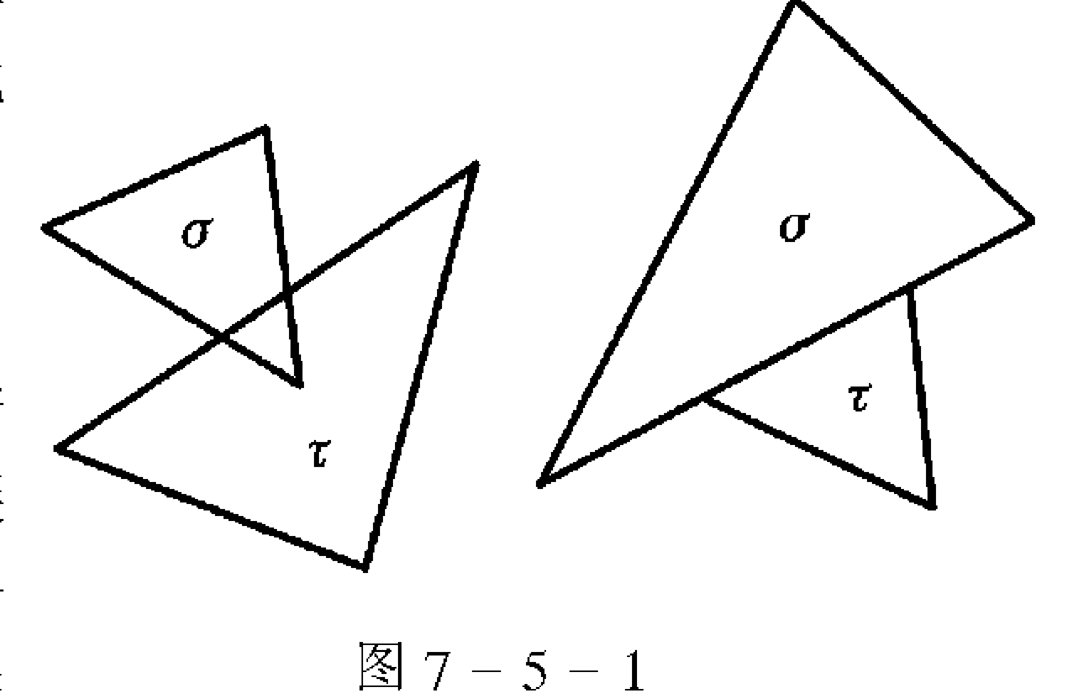



图 7－5－1

现在把 $\partial$ 线性地由一个单形推广到链 $c$ 上，而令

$$\partial c = \sum n_i\partial\sigma^i.$$

由 (6)，$\partial\sigma^i$ 其实已经是一个链，但是维数低 1，所以我们得到链群上的一个映射序列，记作 $(C, \partial)$：

$$(C, \partial): \cdots \xrightarrow{\partial} C_k \xrightarrow{\partial} C_{k-1} \xrightarrow{\partial} \cdots \xrightarrow{\partial} C_0 \xrightarrow{\partial} 0. \tag{8}$$

最后一项为 0 是因为 $C_0$ 是由复形 $K$ 中之一切单形的顶点所成的形式和 (7) 之集合，顶点 $x_i$ 没有边缘，即边缘为空集，我们用 0 表示空集之形式和.

**定理 1**　$\partial^2 = 0$.

**证**　我们只考虑一个单形 $\sigma=\langle x_0,x_1,\cdots,x_k\rangle$，则

$$\partial\sigma=\sum(-1)^i\langle x_0,\cdots,\hat{x}_i,\cdots,x_k\rangle.$$

再施行$\partial$时注意到$\partial\sigma$已是一个链，而$\partial$应施加于其每一项上. 当$\partial$施于其一项$\langle x_0,\cdots,\hat{x}_i,\cdots,x_k\rangle$时应再删去一个顶点 $x_j$. 如果 $j<i$，则删去的是第 $j+1$ 个顶点，应该加一个因子$(-1)^j$；如果 $j>i$，则因 $x_i$ 本已删去，再删的是第 $j$ 个顶点，因此现在添加的因子是$(-1)^{j-1}$. 总之我们有

$$\begin{aligned}\partial^2\sigma=\partial\partial\sigma&=\sum_{j<i}(-1)^{i+j}\langle x_0,\cdots,\hat{x}_j,\cdots,\hat{x}_i,\cdots,x_k\rangle\\&+\sum_{j>i}(-1)^{i+j-1}\langle x_0,\cdots,\hat{x}_i,\cdots,\hat{x}_j,\cdots,x_k\rangle=0.\end{aligned}$$

最后一式的来源是将后一项的 $i,j$ 对调即知它与第一个和式相消. 证毕.

定理1是一个非常重要的定理，它与外微分算子的基本公式 $d^2=0$ 有深刻的对偶关系. 也正因为如此，我们才把可以写为 $d\sigma$ 的微分形式 $\omega=d\sigma$ 称为“上”边缘，英文名词为 co-boundary. 在数学的这一部分中，一个限制词加上了“co-”这样的字头以后总是意味着有某种对偶性. 但其译法不一. 有时称为“协”，如“协变”covariant，有时译为“余”，如“余切丛”cotangent boundle，有时译为“上”，如现在我们讲的“上同调”cohomology.

我们用(6)式来定义$\partial$，其中有因子$(-1)^i$，其目的就是为了证明这个定理.

这样，例如对2单形$\langle x_0,x_1,x_2\rangle$，我们有

$$\partial\langle x_0,x_1,x_2\rangle=\langle x_1,x_2\rangle-\langle x_0,x_2\rangle+\langle x_0,x_1\rangle.$$

注意第二项，$\langle x_0,x_2\rangle$ 是由 $x_0$ 到 $x_2$ 的线段，现在要把它改成$\langle x_2,x_0\rangle$，即由 $x_2$ 到 $x_0$ 的线段，这样做才能使一个单形的定向与它的各个面的定向协调起来(图7－5－2).

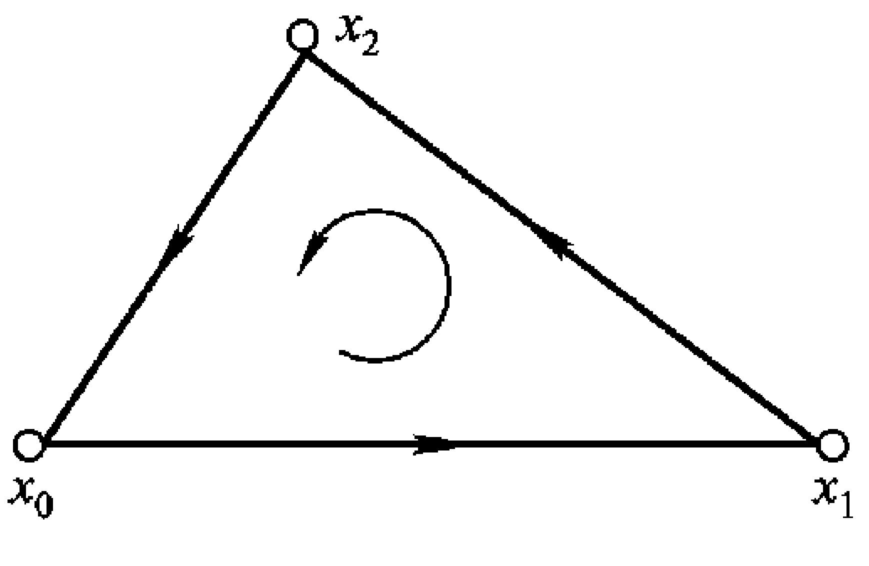



图7－5－2

在以前作积分时，我们要把积分和中的“矩形”(第四章中通称为“区间”)细分. 现在既然用单形来取代矩形，当然就要考虑单形的剖分. 我们总是用一种特殊的剖分称为单形的重心重分(barycentric subdivision). “重心”是一个从初等几何中就熟知的概念，重应该读如轻重的重. “重分”则是再分的意思，所以重分的“重”应该读如“轻舟已过万重山”的“重”.

一个 $k$ 单形$\sigma_k=\langle x_0,x_1,\cdots,x_k\rangle$(其中 $x_i$ 表示$n$ 维线性空间$V$之一点，因此是一个 $n$ 维向量)的重心即

$$b(\sigma_k)=\frac{1}{k+1}\sum_{i=0}^{k}x_i.\tag{9}$$

它仍是一个 $n$ 维向量，因此仍是 $V$ 中一点，但因 $\sigma_k$ 是凸的，(9)中的系数$\dfrac{1}{k+1}>0$，且$\sum\limits_{i=0}^{k}\dfrac{1}{k+1}=1$，所以 $b(\sigma_k)$ 在 $\sigma_k$ 内部，我们熟知的线段中点、三角形重心均是其特例.

**定义2**　$\sigma_k$ 之重心重分$D(\sigma_k)$即将 $\sigma_k$ 分成一切可能的以下形状的一组$k$ 单形，其顶点为$\langle b(\tau_0),b(\tau_1),\cdots,b(\tau_k)\rangle$，这里 $\tau_i$ 为$\sigma_k$ 的一个$i$ 维面，$b(\tau_i)$即其重心，而且 $\tau_i\prec\tau_{i+1}$.

**注意**　$\tau_0$ 是 $\sigma_k$ 的0维面即一个顶点，设为 $x_{i_0}$，$\tau_1$ 既是1维面，而 $\tau_0\prec\tau_1$，所以 $\tau_1$ 必为某个线段$[x_{i_0},x_{i_1}]$，而$b(\tau_1)=\dfrac{1}{2}(x_{i_0}+x_{i_1})$，同样，$b(\tau_k)=\dfrac{1}{k+1}\sum\limits_{j=0}^{k}x_{i_j}$. 但是从上面的作法可

知，$\tau_i$ 的顶点必不能尽取 $\tau_{i+1}$ 的顶点，所以 $\{x_{i_j}\}$ 一定是 $\{x_0,x_1,\cdots,x_k\}$ 的一个重新排列，而 $b(\tau_k)=\frac{1}{k+1}\sum_{i=0}^{k}x_i$ 就是 $b(\sigma_k)$. 当然，我们应该来证明 $\langle b(\tau_0),b(\tau_1),\cdots,b(\tau_k)\rangle$ 确为一个单形. 为此，只要证明它的顶点为仿射无关的即可. 这是容易的. 为方便起见，我们令 $\tau_j=\langle x_0,\cdots,x_j\rangle$，则

$$\sum_{i=0}^{k}\lambda_i b(\tau_i)=\sum_{i=0}^{k}\frac{\lambda_i}{i+1}\sum_{j=0}^{i}x_j=\left(\lambda_0+\frac{\lambda_1}{2}+\cdots+\frac{\lambda_k}{k+1}\right)x_0$$
$$+\left(\frac{\lambda_1}{2}+\cdots+\frac{\lambda_k}{k+1}\right)x_1+\cdots+\frac{\lambda_k}{k+1}x_k.$$

若它等于0，由 $\{x_0,x_1,\cdots,x_k\}$ 之仿射无关性，立即有 $\lambda_k=0$，依次代入前面各项又依次有 $\lambda_{k-1}=\lambda_{k-2}=\cdots=\lambda_0=0$.

还要注意，$\sigma_k$ 之重心重分 $D(\sigma_k)$ 一定成一复形，即它的各个小的 $k$ 维单形 $\langle b(\tau_0),b(\tau_1),\cdots,b(\tau_h)\rangle$ 均规则地相处. 这在直观上是很清楚的，但是应该要证明. 由于我们在这里并不是想要讨论代数拓扑学，所以不停留在这个问题上了. 但有一点应该提及，即在重心重分以后所得到的小的单形的定向应如何确定？从形式上看，每一个小单形 $\langle b(\tau_0),b(\tau_1),\cdots,b(\tau_k)\rangle$ 中 $\tau_i<\tau_{i+1}$，所以这些顶点已经有了一个自然的定向，但是这个定向与原来的单形之定向却不一定协调. 例如，$b(\tau_0)$ 一定是某个顶点 $x_{i_0}$，$b(\tau_1)$ 应是线段 $(x_{i_0},x_{i_1})$ 的中点，但是在原单形中 $(x_{i_0},x_{i_1})$ 的定向不一定是由 $x_{i_0}$ 到 $x_{i_1}$ 而可能是相反，例如图 7－5－3之左. 现在我们要调整如下，我们已说过，$\{i_0,i_1,\cdots,i_k\}$ 一定是 $\{0,1,\cdots,k\}$ 的一个排列 $\pi$，于是我们规定小单形 $\langle b(\tau_0),b(\tau_1),\cdots,b(\tau_k)\rangle$ 的定向应取为 $\operatorname{sgn}\pi\langle b(\tau_0),b(\tau_1),\cdots,b(\tau_k)\rangle$，图 7－5－3右就是调整以后的定向.

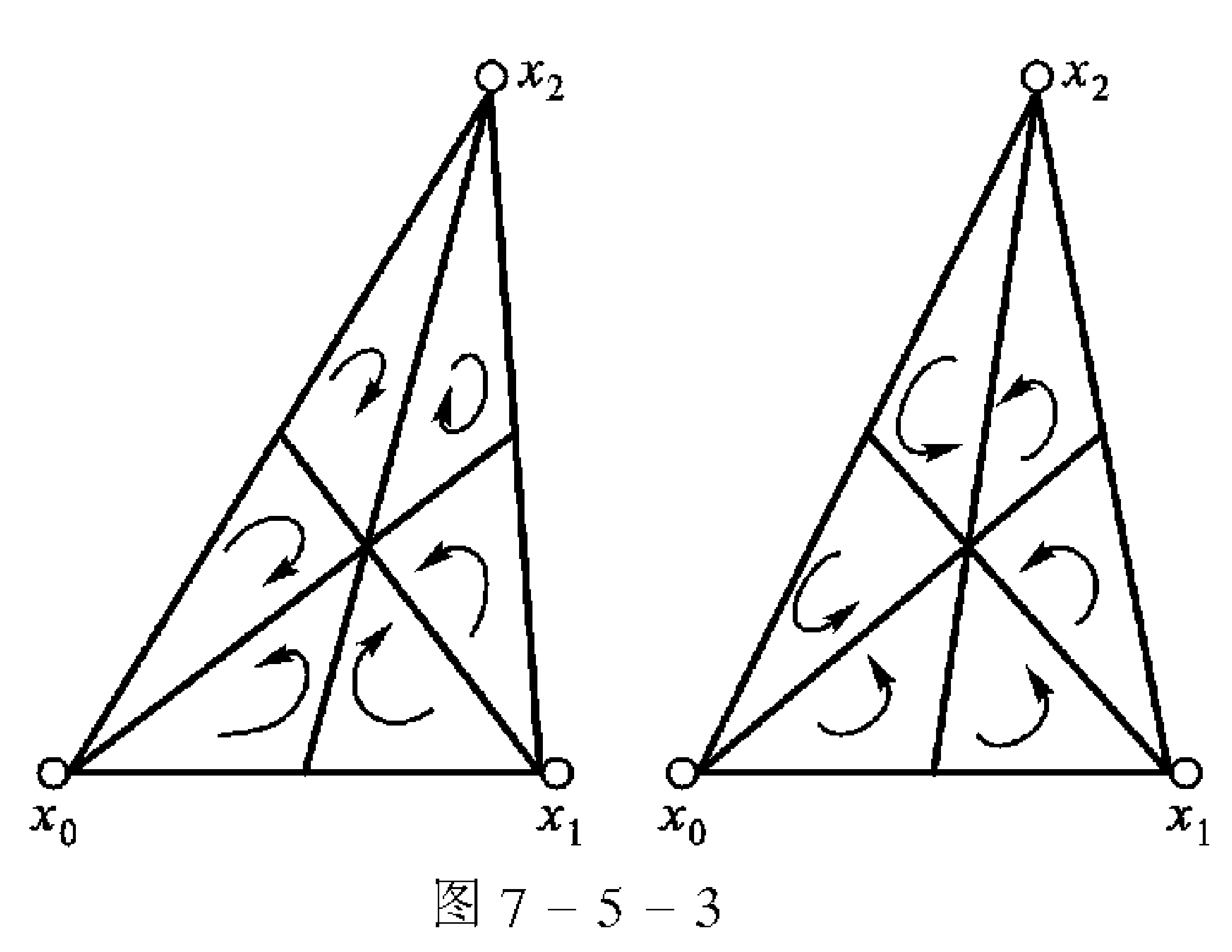



图 7－5－3

我们当然要对 $\sigma$ 反复作重心重分，例如作出 $D^m(\sigma)=D\cdot D\cdot\cdots\cdot D(\sigma)$（共 $m$ 重），而看 $m\to\infty$ 时它是否会缩为一点. 这时就要问了，在 $V$ 上我们并未赋以度量，怎样来考虑极限呢？事实上，$V\cong\mathbf{R}^n$，在 $\mathbf{R}^n$ 中当然可以作出极限，当然可以给出度量（见第六章）而不必要有欧氏度量（这一点与体积不同，要定义体积却是需要欧氏度量的）. 而且第六章中还指出了，$\mathbf{R}^n$ 中各种度量均是等价的. 例如两个点 $x_1=(x_1^1,\cdots,x_1^n)$ 与 $x_2=(x_2^1,\cdots,x_2^n)$ 之间，用 $\sum_i|x_1^i-x_2^i|$，

$\sup_i |x_1^i - x_2^i|$，$\left[\sum_i |x_1^i - x_2^i|^p\right]^{\frac{1}{p}}$ $(p > 1)$ 作距离均无不可．所以下面我们直接写 $|x - y|$ 表示二点的距离，于是令 $\sigma = \langle x_0, x_1, \cdots, x_k \rangle$ 为一个单形，$x, y \in \sigma$，于是 $y = \sum_{i=0}^{k} \lambda_i x_i$，$0 \leqslant \lambda_i \leqslant 1$，$\sum \lambda_i = 1$，所以

$$|x - y| \leqslant \sum_{i=0}^{k} \lambda_i |x_i - x| \leqslant \sup_i |x_i - x| \leqslant \sup_{i,j} |x_i - x_j|.$$

这里我们再把 $x$ 写为 $x = \sum_{j=0}^{k} \mu_j x_j$，然后再用上面的推理．作重心重分后，我们只要讨论 $|b(\tau) - b(\mu)|$ 即可，于是设 $\tau \prec \mu$，$b(\tau)$ 是一个 $l$ 维面的重心，$b(\mu)$ 是一个 $m$ 维面的重心，$l < m \leqslant k$，于是

$$b(\tau) = \frac{1}{l+1} \sum_{j=0}^{l} x_{i_j}, \quad b(\mu) = \frac{1}{m+1} \sum_{j'=0}^{m} x_{i_{j'}}.$$

为了避免符号过于冗长，$x_{i_j}$，$x_{i_{j'}}$ 就写成 $x_j$，$x_{j'}$，这样利用单形为一凸集，即知

$$\left| b(\tau) - b(\mu) \right| \leqslant \sup_{j \leqslant l} \left| x_j - b(\mu) \right| \leqslant \sup_{j \leqslant l} \left| x_j - \frac{1}{m+1} \sum_{j'=0}^{m} x_{j'} \right|$$

$$\leqslant \sup_{j \leqslant l} \left| \frac{1}{m+1} \sum_{j'=0}^{m} (x_j - x_{j'}) \right| \leqslant \frac{m}{m+1} \sup_{j,j'} |x_j - x_{j'}|$$

$$\leqslant \frac{k}{k+1} (\sigma \text{ 之直径}).$$

由此可知，当无限地作重心重分后，单形 $\sigma$ 将缩为一点．

有了这些预备知识以后，我们可以开始讨论积分了．我们首先考虑一个 $k$ 形式 $\omega = a(x) \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k$ 在一个 $k$ 单形 $\sigma_k = \langle x_0, x_1, \cdots, x_k \rangle$ 上的积分．这里遇到了关键问题，即如何处理体积．如果在 $\mathbf{R}^n$ 上已有欧氏结构，而且不妨即设 $(x^1, \cdots, x^k)$ 为笛卡尔坐标，则切空间有一个 o.n. 基底 $\{e_1, \cdots, e_k\} = \left\{\frac{\partial}{\partial x^1}, \cdots, \frac{\partial}{\partial x^k}\right\}$．$\{\mathrm{d}x^1, \cdots, \mathrm{d}x^k\}$ 则是余切空间与它对偶的 o.n. 基底．如果记 $\sigma_k$ 之棱为切向量 $\{v_1, \cdots, v_k\}$（实际上例如 $v_1 = x_1 - x_0$），而且

$$v_i = A_i^j \frac{\partial}{\partial x^j} \qquad (\text{注意 } A_i^j \text{ 为实常数}).$$

于是 $\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k$ 作为 $k$ 阶反对称协变张量，易知

$$\langle \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k; v_1, \cdots, v_k \rangle = \det(A_i^j) \left\langle \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k; \frac{\partial}{\partial x^{j_1}}, \cdots, \frac{\partial}{\partial x^{j_k}} \right\rangle$$

$$= \det(A_i^j) = k! \operatorname{vol}(\sigma_k).$$

今取 $\omega$ 之系数 $a(x)$ 在 $\sigma_k$ 的重心 $b_k$ 处的值 $a(b_k)$，并且作出第一个"黎曼和"．

$$S_1(\omega, \sigma_k) = \frac{1}{k!} \langle a(b_k) \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k; v_1^{(1)}, \cdots, v_k^{(1)} \rangle = a(b_k) \operatorname{vol}(\sigma_k), \tag{10}$$

这里 $v_i^{(1)}$ 就是 $v_i$．(10) 式前的因子 $\frac{1}{k!}$ 来自：$\det(A_i^j)$ 是 $k$ 维平行体的有符号的体积．乘上 $\frac{1}{k!}$ 后才得到单形 $\sigma_k$ 的体积 $\operatorname{vol}(\sigma_k)$．

往下的作法就很清楚了. 作 $\sigma_k$ 的二次重分 $D^2(\sigma_k)$. 它是一个链，由具有适当符号的 $k$ 单形 $\sigma_{k,i}^2$ 构成. 对每个这样的单形，仿照 (10) 作出

$$S_2(\omega,\sigma^{(2)}) = \frac{1}{k!}\sum_i \langle a(b_{k,i}^{(2)})\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k; v_1^{(2)},\cdots,v_k^{(2)}\rangle,$$
$$= \sum_i a(b_{k,i}^{(2)})\mathrm{vol}(\sigma_{k,i}^{(2)}).$$

然后再作出

$$S_m(\omega,\sigma^{(m)}) = \frac{1}{k!}\sum_i \langle a(b_{k,i}^{(m)})\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k;$$
$$v_1^{(m)},\cdots,v_k^{(m)}\rangle = \sum_i a(b_{k,i}^{(m)})\mathrm{vol}(\sigma_{k,i}^{(m)}). \tag{11}$$

我们考虑的是 $\mathbf{R}^k$ 上的光滑微分形式，所以应该看到 $a(x)$ 在有界闭集 $\sigma_k$ 上是连续的，因而是黎曼可积的，因此当 $m\to\infty$ 时，$\lim\limits_{m\to\infty} S_m(\omega,\sigma^{(m)}) = \int_{\sigma_k} a(x)\mathrm{d}x^1\cdots\mathrm{d}x^k$ 是存在的. 所以，我们有理由给出

**定义 3** 若 $\omega = a(x)\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k$ 在 $\mathbf{R}^k$ 中包含某个 $k$ 维单形 $\sigma_k$ 的区域中是光滑的 $k$ 形式，则定义

$$\int_{\sigma_k}\omega = \int_{\sigma_k} a(x)\mathrm{d}x^1\cdots\mathrm{d}x^k. \tag{12}$$

对这个定义应该作一些说明. 在定义 3 前的一大段文字并没有证明什么，而只是解释为什么要这样定义. 古典的积分定义的主要思想，如果给以不严格的表述，就是：把一个函数 $a(x)$，乘上“无穷小的体积”$\mathrm{d}x^1\cdots\mathrm{d}x^k$，再加起来. 现在则是，把一个微分形式 $\omega$ 中的“体积元素”$\mathrm{d}x^1 \wedge\cdots\wedge \mathrm{d}x^k$ 作用到“无穷小单形”$\sigma_{k,i}^{(m)}$ 上并乘以 $a(b_{k,i}^{(m)})$ 然后再加起来，二者实在是很相似的. 问题在于 $\mathbf{R}^k$ 上在没有给出度量以前，是不能定义体积的. $\mathbf{R}^k$ 上当然可以给出度量，而且可以给出许多种：每一个度量就是一个正定的 $k$ 阶方阵 $(g_{ij})$. 上面的讲法则是指定用单位矩阵 $I$ 作为 $(g_{ij})$. 当然会问，如果给出其它度量又当如何？结果当然会有不同，下面讲到函数（而不是微分形式）在黎曼流形（上面当然已经有了度量）上的积分时会要回答这个问题. 可以说其结果无非相当于对系数 $a(x)$（即“被积函数”）作一个修正而已. 这时整个理论不会有本质的改变. 但它究竟告诉我们，微分形式的积分本质地依赖于体积概念. 体积是可以“定义”的：用不同定义的体积将得到不同的积分值. 真正的问题是：$\mathbf{R}^k$ 上本来没有体积，要先给出体积定义以后，才能用这样算出的黎曼积分 $\int_\sigma a(x)\mathrm{d}x^1\cdots\mathrm{d}x^k$ 作为 $\int_\sigma\omega$ 之定义.

我们的作法当然有些偷巧. 既然知道 $\mathrm{d}x^1 \wedge\cdots\wedge \mathrm{d}x^k$ 作用到 $(v_1,\cdots,v_k)$ 上以后可以得到一个“很像”体积的东西，何不以它为基础建立一套测度和积分理论而要将它归结为通常的黎曼积分呢？如果那样做就太麻烦了. 回想一下古典的黎曼积分定义是很麻烦的：一维情况好说（现在几乎一定要考虑高维问题），在高维时，我们需要把积分区域分解成若尔当可测的小区域（而什么条件下一个区域才是若尔当可测又是麻烦事），现在我们只分划为单形，而单形又只允许作重心重分，这当然容易多了. 特别是把它化成了黎曼积分就不必操心证明这个“很像”体积的东西是否具有有限或可数可加性了. 可是以单形及其重心重分为基础来建立积分理论真正的目的是为了

突出积分的组合学的侧面,表现在下面的斯托克斯定理上.

根据同样的"偷巧"的办法,有许多关于积分必须要证明的结果将都在其较一般的情况下,即流形上的积分中,一并证明.

最后还要提一下,我们的 $\mathrm{vol}(\sigma_k)$ 是有符号的:若将 $\sigma_k$ 的定向反转,成为 $-\sigma_k$ 则 $\det(A_j^i)$ 将要反号,于是

$$\int_{-\sigma_k}\omega = -\int_{\sigma_k}\omega. \tag{13}$$

这可能是微分形式的积分与通常函数的积分最重要的区别.也很清楚,若有一个由 $k$ 单形 $\sigma_k^{(i)}$,$i=1,2,\cdots,l$,($l$ 为有限整数)所成的链

$$c=\sum_{i=1}^{l}\lambda_i\sigma_k^{(i)},\lambda_i\in\mathbf{Z} \tag{14}$$

(这里 $\mathbf{Z}$ 指整数集),则可以在链上求积分而有

$$\int_c\omega=\sum_{i=1}^{l}\lambda_i\int_{\sigma_k^{(i)}}\omega. \tag{15}$$

**2. 斯托克斯定理** 许多数学家认为斯托克斯定理是整个数学中最深刻的定理之一.其实,在19世纪中叶有许多类似的定理都差不多同时被人提出,例如高斯、格林的定理.对之有贡献的数学家与数学物理学家还可以举出许多.它之所以重要在于它几乎是自明的,而这又依赖于彻底地弄清楚它所涉及的概念.同时它有十分重要的推论.由于它十分重要,我们宁可多花一些篇幅,在 $\mathbf{R}^n$ 上与在一般的微分流形上各讲一次,这样可以更清楚地理解它所涉及的概念.

**斯托克斯定理** 设 $\sigma$ 为一有向 $k$ 维单形,$\omega$ 为 $\sigma$ 上的光滑的 $(k-1)$ 微分形式,则

$$\int_\sigma \mathrm{d}\omega=\int_{\partial\sigma}\omega. \tag{16}$$

**证** 我们最终将在一个微分流形上证明类似定理,那时我们也将给出其准确的提法.所以现在我们仍只限于说明,从组合学的角度去看,它究竟是怎么回事.

上面我们说过,若 $\sigma=\langle x_0,x_1,\cdots,x_k\rangle$ 则

$$\partial\sigma=\sum_{i=0}^{1}(-1)^i\langle x_0,\cdots,\hat{x}_i,\cdots,x_k\rangle,$$

我们看左、右双方的黎曼和,而且首先是 $\sigma$ 未作重心重分的情况.对于(16)的左方,它是

$$S_1(\mathrm{d}\omega,\sigma)=\frac{1}{k!}\langle \mathrm{d}\omega\mid_{x=b(\sigma)};v_1,\cdots,v_k\rangle.$$

但是上一节中的(21)式(我们只对 $\omega$ 为1微分形式的情况证明了它)告诉我们

$$\langle \mathrm{d}\omega\mid_{x=b(\sigma)};v_1,\cdots,v_k\rangle=\sum_{i=1}^{k}(-1)^{i-1}D_{v_i}\langle\omega;v_1,\cdots,\hat{v}_i,\cdots,v_k\rangle\mid_{x=b(\sigma)}+\text{其它项}.$$

这里"其它项"形如 $\langle\omega(b(\sigma_1));[v_i,v_j],v_1,\cdots,\hat{v}_i,\cdots,\hat{v}_j,\cdots,v_h\rangle$,而且 $D_{v_i}$ 就表示在 $v_i$ 方向的方向导数,亦即(9)式中的 $v_i=A_i^j\dfrac{\partial}{\partial x^j}$.但是 $v_i$ 都是常系数微分算子,因此互相间都可以交换:$[v_i,v_j]=0$,因此"其它项"是不会出现的.故

$$S_1(\mathrm{d}\omega,\sigma)=\frac{1}{k!}\sum_{j=1}^{k}(-1)^{j-1}D_{v_j}\langle\omega(b(\sigma));v_1,\cdots,\hat{v}_j,\cdots v_k\rangle.\tag{17}$$

再看(16)的右方,相应于它的黎曼和是

$$S_1(\omega,\partial\sigma)=\frac{1}{(k-1)!}\sum_{j=0}^{k}(-1)^jS_1(\omega,\sigma_j).$$

$\sigma_j=\langle x_0,\cdots,\hat{x}_j,\cdots,x_k\rangle$,当 $j>0$ 时它是由 $\{v_1,\cdots,\hat{v}_j,\cdots,v_k\}$ 构成的 $(k-1)$ 单形.但是当 $j=0$ 时 $\sigma_0=\langle x_1,\cdots,x_k\rangle$ 的棱是 $\{w_i=x_i-x_1=v_i-v_1\}$, $i=2,\cdots,k$.因此在 $S_1(\omega,\sigma_0)$ 中 $\omega$ 是作用于 $w_2\wedge\cdots\wedge w_k=(v_2-v_1)\wedge\cdots\wedge(v_k-v_1)$ 的.把 $w_2\wedge\cdots\wedge w_k$ 按 $v_1$ 展开,其第一项不含 $v_1$ 而是 $v_2\wedge\cdots\wedge v_k$.含 $v_1$ 的"一次幂"者是在 $(v_2,\cdots,v_j,\cdots,v_k)$ 中以 $-v_1$ 代替 $v_j$,其它项则至少含两个 $v_1$ 作为因子,外微分形式 $\omega$ 既是反对称协变张量,则它作用于含两个(或更多)相同因子 $v_1$ 的项上自然为 0,因此

$$\begin{aligned}S_1(\omega,\partial\sigma)&=\frac{1}{(k-1)!}(-1)^0S_1(\omega(b_0),\sigma_0)+\frac{1}{(k-1)!}\sum_{j=1}^{k}\langle\omega(b_j),\sigma_j\rangle\\&=\frac{1}{(k-1)!}\sum_{j=1}^{k}(-1)^j\langle\omega(b_j),v_1,\cdots,\hat{v}_j,\cdots,v_k\rangle+\\&\quad\frac{1}{(k-1)!}\sum_{j=1}^{k}(-1)^{j-1}\langle\omega(b_0),v_1,\cdots,\hat{v}_j,\cdots,v_h\rangle.\end{aligned}\tag{18}$$

这两项不同在于,第二项中是在 $b_0$ 处取 $\omega$ 之值,而第一项则是在 $\sigma$ 之第 $j$ 个面 $\sigma_j$ 之重心 $b_j$ 处取 $\omega$ 之值.所以

$$S_1(\omega,\partial\sigma)=\frac{1}{(k-1)!}\sum_{j=1}^{k}(-1)^{j-1}\langle[\omega(b_0)-\omega(b_j)],v_1,\cdots,\hat{v}_j,\cdots,v_h\rangle.$$

下面是关键的一步.应用中值定理于 $\omega(b_0)-\omega(b_j)$.在微分学一章中我们就指出过,对于多元函数,中值定理一般不成立,而只能用一个不等式来代替.现在注意单形 $\sigma$ 对其任何一点都是星形的,因此连接 $b_0$ 与 $b_j$ 的线段 $\{tb_0+(1-t)b_j,0\leqslant t\leqslant 1\}$ 恒在 $\sigma$ 内,而 $\omega$ 在这个线段上确实是单变量 $t$ 的函数,因此可以应用中值定理(更准确地说是应用专用于一元函数的拉格朗日公式)而知在此线段上一定可以找到一点 $c_j$ 使

$$\omega(b_0)-\omega(b_j)=\mathrm{D}_{v_j}(c_j)(b_0-b_j).$$

这里 $\mathrm{D}_{v_j}$ 是沿 $(b_0,b_j)$ 线段的方向导数,而因

$$b_0-b_j=\frac{1}{k}\Big(\sum_{i\neq 0}x_i-\sum_{i\neq j}x_i\Big)=\frac{1}{k}(x_j-x_0)=\frac{1}{k}v_j,$$

所以就是 $v_j$ 方向的向量,从而

$$\omega(b_0)-\omega(b_j)=\frac{1}{k}\mathrm{D}_{v_j}(c_j)v_j.$$

代入(18)即有

$$S_1(\omega,\partial\sigma)=\frac{1}{k!}\sum_{j=1}^{k}(-1)^{j-1}\mathrm{D}_{v_j}\langle\omega(c_j),v_1,\cdots,\hat{v}_j,\cdots,v_k\rangle.\tag{19}$$

把(17)与(19)比较,其差别一是在 $c_j$ 处求值,一是在 $b(\sigma)$ 处求值.但在经过多次重心重分后 $b(\sigma)$ 与 $c_j$ 会任意接近,而且若以重心重分后 $k$ 维单形之直径作为无穷小之阶的标准,则每一项

之差均为高阶无穷小量.按照黎曼积分的标准作法,其极限又会相同.由此即得斯托克斯定理之证.

认真分析这个证明,首先我们应看到这个定理是微积分基本定理的推广.因为若 $\sigma$ 是 1 维单形即区间 $[x_0, x_1]$ 时,$\omega$ 应是 0 形式,即一个 $C^\infty$ 函数 $f(x)$,$\mathrm{d}\omega = f'(x)\mathrm{d}x$,而 $\omega$ 在 $\partial\sigma = \{x_1\} - \{x_0\}$ 上的"积分"现在成为 $f(x_1) - f(x_0)$,斯托克斯定理就成了

$$\int_{x_0}^{x_1} f'(x)\mathrm{d}x = f(x_1) - f(x_0).$$

当然读者会说,这里对 $f(x)$ 光滑性要求过高.整个第四章可以说正是围绕可积性而来,那么何必那么大张旗鼓地讲斯托克斯定理是它的推广呢?大凡一个重要理论总有许多不同的侧面.以定积分为例,它的积分区域构造极为简单,可是古代不明白什么叫可积性.到柯西解决了连续函数的积分问题.以后,不连续性问题提到议事日程上来,由黎曼、达布终于又提出了可积性的理论.后来出现了勒贝格的理论,我们在第四章中正是这样展开这个理论的.可是这只是问题的一个侧面.当我们看到,特别是在物理学中的曲线积分与曲面积分,其积分区域的几何构造(准确些说是其拓扑特性)变得十分复杂时,就发现微分流形上的微分形式之积分成了不可避免的对象.这时,研究积分的几何、拓扑侧面就自然凸现出来.而斯托克斯定理就自然成了人们注意的中心.法国大数学家托姆(R.Thom)说过一段话:"要问我最深刻、最困难而又有具体的不容置疑的物理解释的数学定理是哪一个?对于我,斯托克斯定理当属首选.这与下面的事实有关:外微分是一个十分神秘的观念,它的真正的实质,我相信,还藏在谜中,尽管它的形式定义是如此简单."(转引自 D.Leborgne, Calcul differentiel et géométrie, 1982, 146 页).

在上述证明中最令人注意的是什么?一是中值定理的应用.从我们在第四章 §1 中介绍的达布关于微积分基本定理的证明中,我们就看见这个定理的作用,在讲到泰勒展开式的积分余项时,也是利用 $x$ 点与 $x_0$ 点的连线当 $x$ 与 $x_0$ 充分接近,全部落在所考虑的区域内,因而可以利用积分中值定理(也就是拉格朗日公式).在庞加莱引理和这里则明确提出了区域的星形的概念.它与在拓扑学上重要的可收缩性概念相关.我们要指出,在许多其它重要定理中,也都会以类似形式应用中值定理.二是组合方法的应用.即在本书中这也不是第一次了:达布对微积分基本定理的证明就是利用了一个极其简单的组合学事实:

$$F(b) - F(a) = F(a_N) - F(a_0) = \sum_{i=1}^{N} [F(a_i) - F(a_{i-1})].$$

我们也指出过,分部积分法,无非就是连续形式的阿贝尔求和变换式

$$\sum_{k=1}^{N} (a_k - a_{k-1}) b_k = \sum_{k=0}^{N} (b_k - b_{k+1}) a_k, a_0 = b_{N+1} = 0.$$

现在斯托克斯定理的证明,以及它的基础:单形的边缘与复形的边缘计算公式中的符号的选取,都是使用了组合学的技巧,其目的在于正负相消得到十分简洁的结果.我们在前面说积分学还有另一个侧面即组合学的侧面,就是这个意思.当然,积分学不止是一点组合学,还有"一点"逼近,"一点"极限.其实,这里讲的"一点"也都会成为大文章,下面还要再说.现在我们想告诉读者的是这样一个事实:贯串在整个数学中有一些最基本、最简单的思想和方法.但是它们变化万千,丰富多采.我们学好一门学问,最要下力气的正是用一切方法掌握这些为数不多的基本思想,基本方法,以达到返朴归真.可是没有多年独立的钻研,这是绝做不到的.

**3. 微分形式在流形上的积分以及一般的斯托克斯定理**　本节中我们恒设 $M$ 是一个 $n$ 维可定向微分流形.

微分流形上的积分理论有两个最重要之点：其一是这个理论应该是与坐标无关的.其二是斯托克斯定理.第一点我们在一开始就提到了，它其实就是坐标变换公式，它可以由式(1)，即

$$\int_U \omega = \int_U \varphi^* \Omega = \int_{\varphi(U)} \Omega$$

得出.如果 $U$ 是 $\mathbf{R}^n$ 中的一个复形，$\varphi(U)$ 在 $M$ 上.$M$ 是弯曲的，其上本来没有平直的单形和复形.是否可以设想，$\varphi$ 把复形的结构映射到 $M$ 上，即将 $M$ 分成许多小块，每一个小块 $U$ 是一个 $n$ 维单形 $\sigma$ 之像，而 $\sigma$ 之面，分别映射到 $U$ 的一个低维子流形上?如果可以，就称 $U \subset M$ 是一个 $k$ 维奇异单形.称它为奇异的是因为由 $\sigma$ 到 $U$ 的映射 $\varphi$ 允许很“奇异”，甚至不一定是一对一的.这些奇异单形的形式和 $\sum_i \sigma_i$ 称为奇异链.如果存在这样一个映射，就说 $M$ 可以三角剖分.如果这个映射是 $C^\infty$ 的则说 $M$ 可以 $C^\infty$ 地三角剖分.那么，$M$ 是否可以 $C^\infty$ 三角剖分呢?从20世纪30年代起这就是一个重大问题，因为它关系到能否把组合的方法用于微分流形的研究.这里有惠特尼(Hassler Whitney)的三角剖分定理，指出每一个 $C^\infty$ 微分流形一定有 $C^\infty$ 三角剖分.这个定理的证明太复杂，不妨留待读者有志于更深入地讨论流形的拓扑学时再去研讨.现在我们用另一种方法来讨论，这就是应用单位分解.

在微分流形 $M$ 的定义中我们加上了 $M$ 为豪斯道夫空间并且适合第二可数性公理的要求(§2)，那时我们就指出这样做是为了使用单位分解.其实前面我们已多次这样做了，现在再比较明确地讲一下定义积分的过程，它是应用单位分解方法的一个典型例子.所谓单位分解就是：在上述关于 $M$ 的限制下，若 $M$ 有一个开覆盖，则它必有一个局部有限的子覆盖 $\{U_\alpha\}$，即每一点 $P$ 必有一个邻域 $W$，与最多有限多个 $U_\alpha$(设为 $U_1, \cdots, U_k$)相交：$U_i \cap W \neq \varnothing, i = 1, 2, \cdots, k$.然后可以找到一族 $C_0^\infty(M)$ 函数 $\varphi_\alpha$，称为从属于此覆盖的单位分解，使每一个 $\varphi_\alpha$ 的支集 $\operatorname{supp} \varphi_\alpha$ 必含于某一个 $U_\beta$ 中，$\beta = \beta(\alpha)$(不说是含于 $U_\alpha$ 中，因为 $\{U_\beta\}$ 不一定与 $\{\varphi_\alpha\}$ 用同样的参数标记，而且同一个 $\operatorname{supp} \varphi_\alpha$ 可以含于好几个不同的 $U_\beta$ 中，我们选定其中一个并记为 $\beta(\alpha)$)，这些函数的支集 $\{\operatorname{supp} \varphi_\alpha\}$ 一定是局部有限的，而且

$$0 \leqslant \varphi_\alpha \leqslant 1, \sum_\alpha \varphi_\alpha = 1. \tag{20}$$

后一式要注意，因为有局部有限性的要求，每一点 $P$ 至多含于有限多个 $\operatorname{supp} \varphi_\alpha$ 中，所以上式形状上看来是无限和，其实在 $M$ 之每一点上都是有限和.

现在我们讨论具有紧支集的 $n$ 微分形式 $\omega$ 在 $M$ 上的积分.限制 $\omega$ 具有紧支集的作用如同在黎曼积分中避免反常积分.实际上，令上述 $\{U_\alpha\}$ 是 $M$ 上的局部坐标邻域系，作从属于它的单位分解 $\{\varphi_\alpha\}$，于是由(20)有

$$\omega = \sum_\alpha \varphi_\alpha \omega = \sum \omega_\alpha. \tag{21}$$

而且因为 $\operatorname{supp} \omega$ 为紧，它最多与有限多个 $\operatorname{supp} \varphi_\alpha$ 相交，因此上式确实是一个有限和.其每一项可以写为

$$\omega_\alpha = a_\alpha(x) \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n, \tag{22}$$

$(x^1,\cdots,x^n)$ 是 $U_\alpha$ 中的局部坐标，而 $a_\alpha(x)$ 不妨认为是定义在整个 $\mathbf{R}^n$ 上的，只不过在 $V_\alpha=\varphi_\alpha^{-1}(U_\alpha)\subset\mathbf{R}^n$（$\varphi_\alpha$ 是 $U_\alpha$ 中的坐标映射）之外 $a_\alpha(x)\equiv 0$. 这样 $a_\alpha(x)\in C_0^\infty(\mathbf{R}^n)$，这个函数的黎曼积分没有一点困难.

于是我们得到了 $\mathbf{R}^n$ 上的 $n$ 微分形式，其积分是可以定义的，当然要对 $\mathbf{R}^n$ 赋以典则的，即标准的欧氏度量. 对(21)式之每一项作类似处理后，即可看出，应有

**定义 4** 设 $M$ 是一个可定向的 $n$ 维微分流形，$\omega$ 是 $M$ 上的一个具有紧支集的 $n$ 微分形式，则我们定义 $\omega$ 在 $M$ 上的积分为

$$\int_M\omega=\sum_\alpha\int_M\omega_\alpha=\sum_\alpha\int_{\mathbf{R}^n}a_\alpha(x)\mathrm{d}x^1\cdots\mathrm{d}x^n. \tag{23}$$

与定义3比较，在那里我们并未假设 $\omega$ 具有紧支集，但是积分域是一个单形 $\sigma_k$ 或者是有限个单形所成的链，因而是紧的. 这也是为了避免反常积分那样的困难.

为了说明这个定义的合理性应该回答两个问题：首先，若采用不同的单位分解；其次，若对同一个 $\omega_\alpha$ 采用不同的局部坐标系，所得到的结果是否一致？若不一致，则这个定义是不合理的.

第一个问题很容易回答. 因为若有另一个可能是从属于另一个局部有限覆盖的单位分解 $\{\psi_\beta\}$，$0\leqslant\psi_\beta\leqslant 1$，$1=\sum\limits_\beta\psi_\beta$，则应定义

$$\int_M\omega=\sum_\beta\int_M\psi_\beta\omega.$$

但是

$$\begin{aligned}\int_M\omega&=\sum_\alpha\int_M\varphi_\alpha\omega=\sum_\alpha\Big[\sum_\beta\int_M(\varphi_\alpha\psi_\beta)\omega\Big]\\&=\sum_\beta\Big[\sum_\alpha\int_M(\psi_\beta\varphi_\alpha)\omega\Big]=\sum_\beta\int_M\psi_\beta\Big(\sum_\alpha\varphi_\alpha\omega\Big)\\&=\sum_\beta\int_M\psi_\beta\omega.\end{aligned}$$

所以采用不同的单位分解并不影响定义.

其次，若在积分区域中 $U_\alpha$ 中有两个局部坐标系 $(V,\psi)$ 与 $(W,\tilde{\psi})$，$V=\psi^{-1}(U_\alpha)$，$W=\tilde{\psi}^{-1}(U_\alpha)$，而且局部坐标分别为 $(x^1,\cdots,x^k)$ 与 $(y^1,\cdots,y^k)$，则有微分同胚 $\varphi:V\to W$，$x=\varphi(y)$ 而 $\omega_\alpha$ 有两个表达式

$$\omega_\alpha=a_\alpha(x)\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n=A_\alpha(y)\mathrm{d}y^1\wedge\cdots\wedge\mathrm{d}y^n,$$

$$a_\alpha(x)=A_\alpha(y)\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}.$$

因为我们已设 $M$ 是可定向的，当然可以要求所有的局部坐标的定向是相协调的，因此 $\dfrac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}>0$. 于是

$$\begin{aligned}\int\omega_\alpha&=\int a_\alpha(x)\mathrm{d}x^1\cdots\mathrm{d}x^n=\int A_\alpha(y)\left|\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}\right|\mathrm{d}x^1\cdots\mathrm{d}x^n\\&=\int A_\alpha(y)\mathrm{d}y^1\cdots\mathrm{d}y^n=\int A_\alpha(y)\mathrm{d}y^1\wedge\cdots\wedge\mathrm{d}y^n.\end{aligned}$$

正如本节一开始时就强调指出的，我们要讨论微分形式而不是函数在流形上的积分，正是因为微分形式在局部坐标下的表达式自动地含有我们所需要的变量变换公式，而假设 $M$ 是可定向的就可以保证雅可比行列式中自动地会出现绝对值.

有了定义 4 以后，关于流形 $M$ 上的积分之性质，有

**定理 2**　若以下的 $n$ 微分形式均在可定向 $n$ 维微形 $M$ 上具有紧支集，或 $M$ 为紧流形，则

(1) $\int_M(\lambda_1\omega_1+\lambda_2\omega_2)=\lambda_1\int_M\omega_1+\lambda_2\int_M\omega_2\quad \lambda_1,\lambda_2$ 为实数.

(2) 若将 $M$ 上定向反转，记所得流形为 $-M$，则

$$\int_{-M}\omega=-\int_M\omega. \tag{24}$$

证明很简单，现在均略去，不过要说明，它们对 $\mathbf{R}^n$ 上的积分也都成立.

这里我们要对 (24) 式多说几句. 通常的微积分教本中讲曲线积分与曲面积分时，总分成第一型与第二型两种. 其实所谓"第二型"就是在一维流形（曲线）或二维流形（曲面）上积分. 在那里我们是说函数的积分，现在则应说是微分形式 $P\mathrm{d}x+Q\mathrm{d}y+R\mathrm{d}z$ 或 $P\mathrm{d}y\mathrm{d}z+Q\mathrm{d}z\mathrm{d}x+R\mathrm{d}x\mathrm{d}y$ 的积分（前者是曲线积分，后者是曲面积分）. 不过曲面积分我们是写成 $P\mathrm{d}y\wedge\mathrm{d}z+Q\mathrm{d}z\wedge\mathrm{d}x+R\mathrm{d}x\wedge\mathrm{d}y$. 第二型曲线积分适合 (24) 是人们都明白的：

$$\int_{AB}P\mathrm{d}x+Q\mathrm{d}y+R\mathrm{d}z=-\int_{BA}P\mathrm{d}x+Q\mathrm{d}y+R\mathrm{d}z.$$

对曲面积分我们则说在此曲面的这一侧或另一侧积分，而单侧曲面如默比乌斯带上是否可以积分则不涉及. 其实这里也一样，我们只讨论可定向流形上的积分. 不可定向流形上如何积分是另一个理论的内容. 通常微积分教本中讲曲面的侧时例如是说：在曲面上指定一侧，如果我们头向这一侧站立，则在此曲面上规定逆时针方向为正向；另一侧就是，如果头向下倒立时看见的逆时针方向为正向，所以两侧有相反的定向. 之所以可以这样做是因为我们承认曲面是在三维空间（即其包含空间）中的，所以有头向上、向下之说. 现在的流形 $M$ 则没有包含空间. 所以无所谓"侧". 要规定其中的正向，就要规定好流形之定向. 不过我们有时仍借用"侧"的说法，而把 (24) 式说成是在 $M$ 之不同侧上的积分反号. 至于第一型线积分和曲面积分，只不过是采用弧长或曲面面积 $\mathrm{d}s,\mathrm{d}\sigma$ 作为测度. 它们都是正的，这与前面讲的黎曼或勒贝格多重积分是一样的，所以没有与 (24) 相应的性质. 也可以说，本书前面讲的多重积分是"第一型"的，但是定积分 $\int_a^b f(x)\mathrm{d}x$ 却是第二型的，因为它有与 (24) 一样的性质.

本节一开始我们就指出了微分流形上的积分定义必须是与坐标的选取无关的. 因此，积分中变量变换公式起了根本的作用，其实这个公式是一个更一般的定理的特例. 设有两个 $n$ 维微分流形 $M$ 和 $N$ 以及一个微分同胚 $\varphi:M\to N$，若 $\omega$ 是 $N$ 上的一个具有紧支集的 $n$ 微分形式，则它在 $M$ 上的拉回：$\varphi^*\omega$ 是 $M$ 上的一个 $n$ 微分形式，而且也具有紧支集. 这时，我们有

**定理 3**　若 $\varphi,M,N$ 均如上述，则

$$\int_M\varphi^*\omega=\varepsilon\int_{\varphi(M)}\omega. \tag{25}$$

这里 $\varepsilon=\pm1$ 视 $\varphi$ 为保持定向或反转定向而定.

**证**　在必要时引用上面所说的单位分解，我们可以设 $C=\operatorname{supp}\omega\Subset U$，$U$ 是 $N$ 上一个坐标

邻域，而坐标映射 $f: U \to \mathbf{R}^n$，$f(C) \subset f(U)$ 中的局部坐标为 $(y^1,\cdots,y^n)$，使得

$$\omega = a(y^1,\cdots,y^n)\mathrm{d}y^1 \wedge \cdots \wedge \mathrm{d}y^n.$$

不失一般性，可以设 $\varphi^{-1}(U)$ 在 $M$ 的一个坐标邻域中，其局部坐标是 $(x^1,\cdots,x^n)$. 于是

$$\varphi^* \omega = A(x^1,\cdots,x^n)\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n,$$

而且

$$A(x^1,\cdots,x^n) = a(y^1,\cdots,y^n)\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)} = \varepsilon a(y^1,\cdots,y^n)\left|\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}\right|.$$

所以

$$\int \varphi^* \omega = \int A(x^1,\cdots,x^n)\mathrm{d}x^1\cdots\mathrm{d}x^n = \varepsilon\int a(y^1,\cdots,y^n)\left|\frac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}\right|\mathrm{d}x^1\cdots\mathrm{d}x^n$$

$$= \varepsilon\int a(y^1,\cdots,y^n)\mathrm{d}y^1\cdots\mathrm{d}y^n = \varepsilon\int\omega.$$

必要时应用单位分解就可得到(25).

在上面的证明中，$M$ 与 $N$ 是可定向的是一个重要条件. 因为不然的，$\varepsilon$ 可能在不同坐标邻域上不同. 但是退一步说，还得 $M$ 与 $N$ 都是连通的，否则在不同的连通分支上 $\varepsilon$ 可能各取 $+1$ 或 $-1$ 而不同. 这些都是很容易理解的. 重要的是要明白 $\varphi$ 为微分同胚是一个本质的要求，否则的话 (25) 会不成立，而要改为

$$\int_M \varphi^* \omega = \varepsilon \deg \varphi \int_{\varphi(M)} \omega.$$

这里的 $\deg\varphi$ 是一个整数，称为映射度(deg ree of mapping). 粗略地说，就是 $N$ 被 $\varphi(M)$ 覆盖的次数. 其实在前面讲到微分形式的拉回时，就出现了这个问题，不过我们没有深究.

如果 $M = N$，定理 3 就是积分的变量变换公式.

现在我们要进到流形上的斯托克斯定理的证明. 它是本节最主要的结果. 我们先叙述这个定理，并详细解释其中涉及的概念. 在做完了这一切以后，读者会看到，它几乎是自明的. 其实前面我们已经指出了，这个定理的证明正是依赖于彻底弄明白它所涉及的概念.

**定理 4 (斯托克斯定理)**　设 $M$ 是一个可定向的 $n$ 维微分流形，而且具有光滑的边缘 $\partial M$，若 $\omega$ 是 $M$ 上的具有紧支集的 $(n-1)$ 微分形式，则

$$\int_M \mathrm{d}\omega = \int_{\partial M}\omega. \tag{26}$$

带边的流形在 §2 中详细讨论过. 它就是这样的流形，即 §2 图 7－2－4 中的区图 $\{U_\alpha\}$ 有两类，一类同胚于 $\mathbf{R}^n$ 中的一个开集，另一类则不能与任一 $\mathbf{R}^n$ 开集同胚，但同胚于半空间 $H^n = \{x; x \in \mathbf{R}^n, x_n \geqslant 0\}$ 中的一个开集. $H^n$ 如果作为 $\mathbf{R}^n$ 的子集，借用 $\mathbf{R}^n$ 的拓扑，则以 $x_n = 0$ 上的点为边界点. $H^n$ 中既然含了若干边界点，则作为 $\mathbf{R}^n$ 的子集就不能看作是一个开集. 但是 $H^n$ 还可以赋予 $\mathbf{R}^n$ 的子空间的拓扑，在这个拓扑下，$H^n$ 自身就是一个拓扑空间的全空间，它就应该是一个开集. $x_n = 0$ 上的点在这个拓扑下就不能再看作边界点了. 它们也不可能是外点，因为所谓全空间就是包罗万象，无所谓外. 这样一来，凡讲到 $H^n$ 上的连续函数、光滑函数、连续映射 …… 按我们通常的说法，即按 $\mathbf{R}^n$ 的观点来看，都应该指连续到边、光滑到边 …… 从微积分学的观点来看，一个函数或一个映射在区域内光滑或者光滑到边，有严重的区别，常造成极严重的困难. 但是按 $H^n$ 的观点来看，却都是题中之意，本来就是这么一回事. 例如我们说某一个 $U_\alpha$ 属于第二类，

而 $U_\alpha$ 中有一点 $P$（不能问 $U_\alpha$ 外的点如何，因为流形的定义是讲 $U_\alpha$ 内的点与 $\mathbf{R}^n$ 或 $H^n$ 的一个开集同胚）对应于 $H_n$ 中 $x_n=0$ 上的某一点 $\tilde{P}$，则 $\tilde{P}$ 在 $H_n$ 中的一个邻域（图 7－5－4 的右图中的阴影区域）必与 $M$ 中含 $P$ 的 $U_\alpha$（图 7－5－4 左图中的阴影区域）同胚．而 $\partial M$ 中含 $P$ 的子集 $AB$ 应该同胚于 $H^n$ 中 $x^n=0$ 上含 $\tilde{P}$ 的子集 $\tilde{A}\,\tilde{B}$．于是 $M$ 中的点是否可以分成两类，一类与 $H^n$ 的某一子集中 $x^n=0$ 上的点对应，一类则与 $H^n$ 中适合 $x^n>0$ 之点或与 $\mathbf{R}^n$ 中之点（自然也可以设此点适合 $x^n>0$）相应？关键在于弄明白，决不可能有一个点 $P\in M$，在某个坐标系 $(x^1,\cdots,x^n)$ 下与 $x^n>0$ 的点 $Q$ 相应，而在另一个坐标系 $(y^1,\cdots,y^n)$ 下则与 $y^n=0$ 之点 $\tilde{C}$ 相应（图 7－5－5）．这是因为，如果这样的点存在，则必有图 7－5－5 上的两个开集（右图是 $H^n$ 中的开集）即阴影区域通过迁移函数的 $\psi\circ\varphi^{-1}$ 成为微分同胚：

$$y^i=\Phi^i(x^1,\cdots,x^n).\tag{27}$$

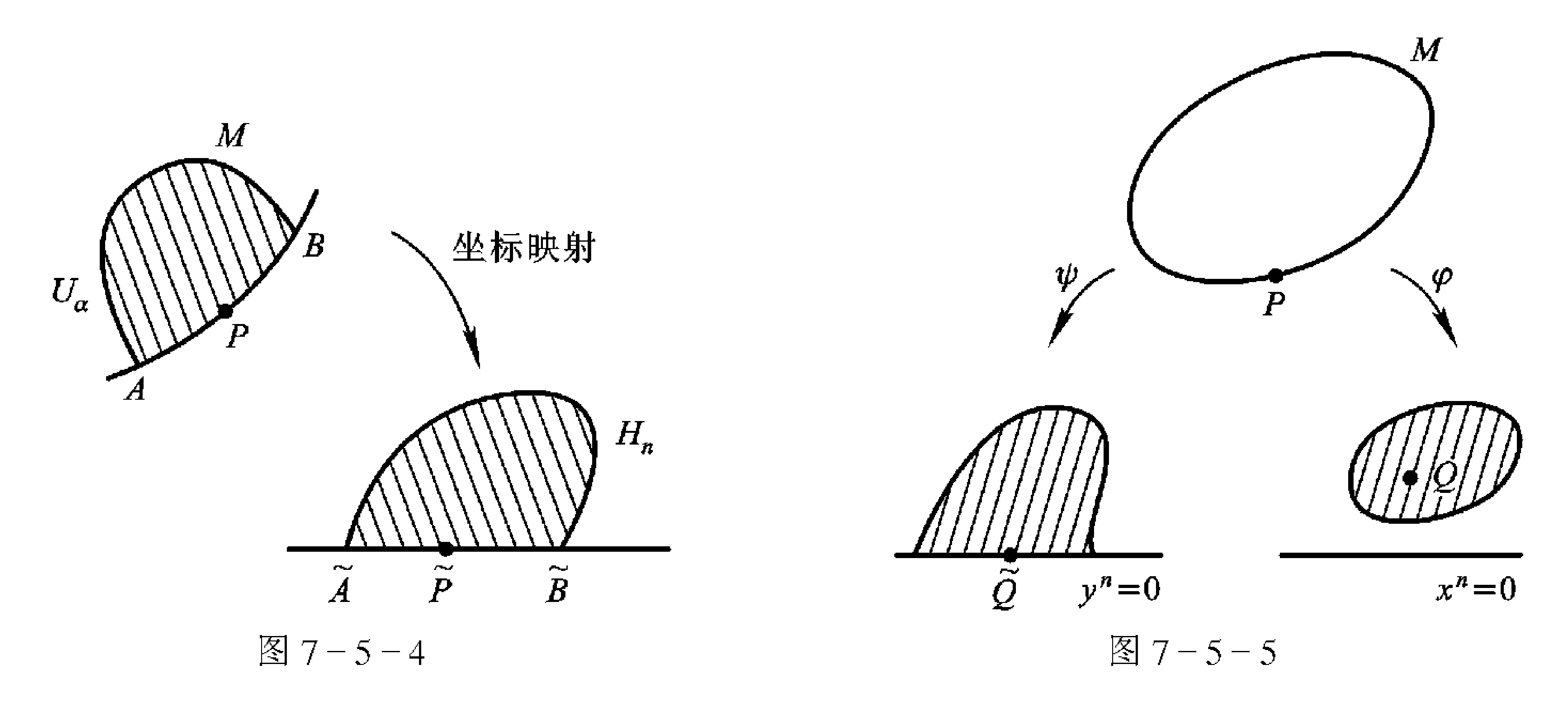



图 7－5－4　　　　　　　　　　　　　　图 7－5－5

而且其雅可比行列式 $\det\dfrac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}\neq 0$ 一直到边．“一直到边”四个字很关键．我们可以假设它 $>0$．如果没有“一直到边”四个字，则只能假设它 $\geqslant 0$．这里不会有 $=0$ 倒不是极限问题，而是因为由微分流形的定义，$\psi\circ\varphi^{-1}$ 既是一直到边的微分同胚，则其逆也是一直到边的微分同胚，所以其逆行列式 $\det\dfrac{\partial(x^1,\cdots,x^n)}{\partial(y^1,\cdots,y^n)}$ 必须一直到 $y^n=0$ 处都连续，这样就必须原行列式在 $y^n=0$ 处不为 0．于是可以对 (27) 在 $Q$ 点应用隐函数定理 —— 这确实是一个 $\mathbf{R}^n$ 中的定理而丝毫不受上面关于 $H^n$ 的拓扑的议论影响 —— 知 $Q$ 附近必有一些点被 (27) 映到 $\tilde{Q}$ 附近 $y^n<0$ 处而与这些迁移函数 $\psi\circ\varphi^{-1}$ 是两个阴影区域间的微分同胚相矛盾．这样一来就看到了 $M$ 中的点有两类：第一类是图册中第一类区图（即 $\mathbf{R}^n$ 中某个开集）内点之像，或是第二类区图（即 $H^n$ 中某个开集）适合 $x^n>0$ 的点之像．第二类则是 $H^n$ 中某开集中适合 $x^n=0$ 之点的像．这两类泾渭分明，绝不混淆．第二类点之集合称为 $M$ 之边缘，记作 $\partial M$．我们在第六章讲拓扑空间中简单地提了一下边缘与边界点之区别，现在就看得比较清楚了．$\partial M$ 并不是一般拓扑空间理论讲的边界点之集合．一是因为现在没有了包含空间的概念，所以没有了外点．二是因为我们 $\partial M$ 之定义并不是它“不是什么什么”而是非常明确的它是什么什么：它是 $H^n$ 中 $x^n=0$ 之像，与一个超平面 $x^n=0$ 有密切关系，所以我们有

**定理 5**　可定向的 $n$ 维带边流形之边缘 $\partial M$ 是一个 $(n-1)$ 维可定向微分流形.

**证**　作 $M$ 的所有第二类区图 $(U_\alpha,\varphi_\alpha)$，记 $V_\alpha=U_\alpha\cap\varphi_\alpha^{-1}(\{x^n=0\})$，则 $\{V_\alpha\}$ 成为 $\partial M$ 的一个覆盖，它与 $\mathbf{R}^{n-1}$ 的开集 $\varphi_\alpha(U_\alpha)\cap\{x^n=0\}$ 显然是同胚的，而且 $\{x^1,\cdots,x^{n-1}\}$ 是它的一个局部坐标系. 现在只要找出不同的 $V_\alpha,V_\beta(V_\alpha\cap V_\beta\neq\varnothing)$ 之间的迁移函数，并证明它们是微分同胚即可. 借用前文的记号，$\varphi_\beta(U_\beta)$ 中的局部坐标为 $(y^1,\cdots,y^n)$，而 $V_\beta=U_\beta\cap\varphi_\beta^{-1}(\{y^n=0\})$. $x^n=0$ 上之点一定被 $\varphi_\beta\circ\varphi_\alpha^{-1}$ 映到 $y^n=0$ 上，这是前面证明了的. 于是这个迁移函数的雅可比行列式就是

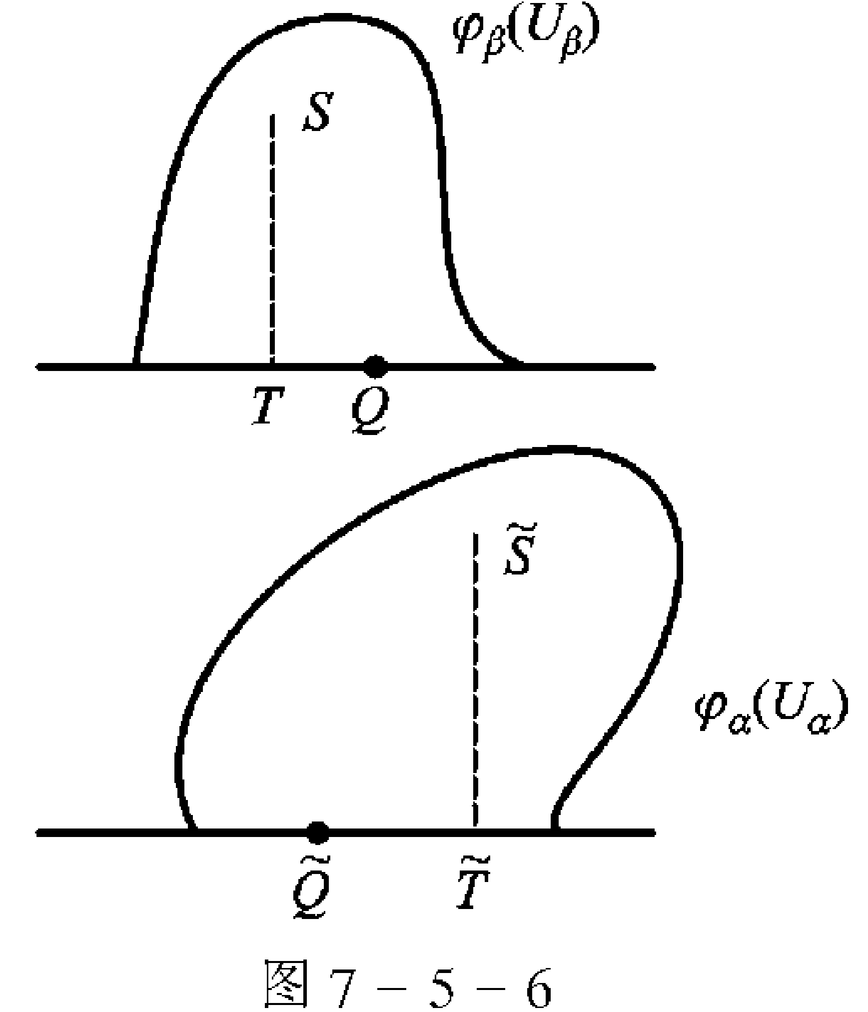


图 7－5－6

$$\begin{vmatrix} \dfrac{\partial y^1}{\partial x^1} & \cdots & \dfrac{\partial y^1}{\partial x^{n-1}} & \dfrac{\partial y^1}{\partial x^n} \\ \\ \dfrac{\partial y^{n-1}}{\partial x^1} & \cdots & \dfrac{\partial y^{n-1}}{\partial x^{n-1}} & \dfrac{\partial y^{n-1}}{\partial x^n} \\ \dfrac{\partial y^n}{\partial x^1} & \cdots & \dfrac{\partial y^n}{\partial x^{n-1}} & \dfrac{\partial y^n}{\partial x^n} \end{vmatrix}$$

左上角的 $(n-1)$ 阶子行列式在 $y^n=0$ 上的限制. 这个限制成为

$$\begin{vmatrix} \dfrac{\partial y^1}{\partial x^1} & \cdots & \dfrac{\partial y^1}{\partial x^{n-1}} & \dfrac{\partial y^1}{\partial x^n} \\ \\ \dfrac{\partial y^{n-1}}{\partial x^1} & \cdots & \dfrac{\partial y^{n-1}}{\partial x^{n-1}} & \dfrac{\partial y^{n-1}}{\partial x^n} \\ 0 & \cdots & 0 & \dfrac{\partial y^n}{\partial x^n} \end{vmatrix}_{\partial M} = \left.\frac{\partial y^n}{\partial x^n}\right|_{\partial M} \begin{vmatrix} \dfrac{\partial y^1}{\partial x^1} & \cdots & \dfrac{\partial y^1}{\partial x^{n-1}} \\ \\ \dfrac{\partial y^{n-1}}{\partial x^n} & \cdots & \dfrac{\partial y^{n-1}}{\partial x^{n-1}} \end{vmatrix}_{\partial M} = \left.\frac{\partial y^n}{\partial x^n}\right|_{\partial M} J.$$

$J$ 就是我们要求的迁移函数雅可比行列式. 今证 $\left.\dfrac{\partial y^n}{\partial x^n}\right|_{\partial M}>0$. 注意前面关于光滑到边的讨论，

$$\left.\frac{\partial y^n}{\partial x^n}\right|_Q=\lim\frac{ST}{\tilde{S}\,\tilde{T}}\geqslant 0$$

(见图 7－5－6). 但是它不可能是 0，否则 $\varphi_\beta\circ\varphi_\alpha^{-1}$ 的雅可比行列式为 0 而没有逆. 因此 $\left.\dfrac{\partial y^n}{\partial x^n}\right|_Q>0$. 因为我们已设 $M$ 是可定向的，所以不妨设 $\dfrac{\partial(y^1,\cdots,y^n)}{\partial(x^1,\cdots,x^n)}>0$ 再加上 $\left.\dfrac{\partial y^n}{\partial x^n}\right|_Q>0$，即得 $J>0$.

由此，我们不但知道了在 $\partial M$ 上 $\{V_\alpha\}$ 构成了一个微分流形的图册，从而 $\partial M$ 是一个 $(n-1)$ 维微分流形，而且 $M$ 之给出相同定向的局部坐标系 $\{(x^1,\cdots,x^n)\}$ 在 $x^n=0$ 上的限制也给出 $\partial M$ 上的一个定向. 因此，$\partial M$ 是有定向的. 定理 5 证毕.

以上我们只不过是解释了定理 4 的陈述中所说 $M$“具有光滑的边缘”是什么意思，现在我们来给出

**斯托克斯定理的证明**　因为 $\omega$ 在 $M$ 上有紧支集，所以它只与有限多个坐标邻域相交. 这些坐标邻域中有一些是第一类的，记为 $U_\alpha^{(1)}$，有一些是第二类的，记为 $U_\beta^{(2)}$. 我们不妨就用

$[\bigcup_\alpha U_\alpha^{(1)}]\bigcup[\bigcup_\beta U_\beta^{(2)}]$代替$M$. 如果用 $V_\beta$ 记 $U_\beta^{(2)}$ 中相应于$H^n$ 之边缘 $x^n=0$ 的点集，则 $\partial M=\bigcup_\beta V_\beta$. 于是作从属于 $M$ 之覆盖$\{U_\alpha^{(1)},U_\beta^{(2)}\}$ 的单位分解$\{\varphi_\alpha^{(1)},\varphi_\beta^{(2)}\}$，有

$$\omega=\sum_\alpha\varphi_\alpha^{(1)}\omega+\sum_\beta\varphi_\beta^{(2)}\omega=\sum_\alpha\omega_\alpha^{(1)}+\sum_\beta\omega_\beta^{(2)},$$

$$\int_M\mathrm{d}\omega=\sum_\alpha\int_M\mathrm{d}\omega_\alpha^{(1)}+\sum_\beta\int_M\mathrm{d}\omega_\beta^{(2)}.$$

先看$\int_M\mathrm{d}\omega_\alpha^{(1)}$. 按定义它可以化为 $U_\alpha^{(1)}$ 在 $\mathbf{R}^n$ 中的像(仍记为 $U_\alpha^{(1)}$) 上的 $\mathbf{R}^n$ 中之积分. 因为 $\mathrm{d}\omega_\alpha^{(1)}$ 之支集不会超出 $\omega_\alpha^{(1)}$ 之支集，所以不妨就设它的积分域是 $\mathbf{R}^n$，而被积函数形如 $\sum_{i,j}\frac{\partial F_j}{\partial x^i}$，$F_j\in C_0^\infty(\mathbf{R}^n)$，但是由微积分的基本定理即有

$$\int_{\mathbf{R}^n}\frac{\partial F_j}{\partial x_i}\mathrm{d}x=0,$$

所以所有的

$$\int_M\mathrm{d}\omega_\alpha^{(1)}=0.$$

再看$\int_M\mathrm{d}\omega_\beta^{(2)}$. 和上面一样的理由，可以设它的积分域是 $H^n$ 的一个很大的单形$\sigma_\beta$，而这个单形有一个面在 $x^n=0$ 上，由于在单形上斯托克斯定理是成立的，所以

$$\int_M\mathrm{d}\omega_\beta^{(2)}=\int_{\sigma_\beta}\mathrm{d}\omega_\beta^{(2)}=\int_{\partial\sigma_\beta}\omega_\beta^{(2)}.$$

但是 $\omega_\beta^{(2)}$ 又是有紧支集的——在 $\mathbf{R}^n$ 的开集中，紧子集离开集的边界点是有一定距离的. 但现在 $\sigma_\beta$ 是在$H^n$ 中，它的支集可以直到 $x^n=0$，即上面讲的 $V_\beta$ 之像. 但我们仍用 $V_\beta$ 表示 supp $\omega_\beta^{(2)}$ 在 $x^n=0$ 上的部分，所以有

$$\int_{\partial\sigma_\beta}\omega_\beta^{(2)}=\int_{V_\beta}\omega_\beta^{(2)}.$$

对 $\beta$ 求和后有

$$\sum_\beta\int_M\mathrm{d}\omega_\beta^{(2)}=\sum_\beta\int_{V_\beta}\omega_\beta^{(2)}=\sum_\beta\int_{\partial M}\omega_\beta^{(2)}.$$

因为 $\omega_\alpha^{(1)}\Big|_{\partial M}=0$，所以综合以上结果有

$$\int_M\mathrm{d}\omega=\int_{\partial M}\omega.$$

定理证毕.

以上我们是将定理归结到单形的情况. 上一段我们已讲了如何在单形边缘上规定定向以便使 $\partial^2=0$，所以在以上的结果中 $\partial M$ 上自动有一个定向使斯托克斯定理成立. 这个使得斯托克斯定理得以成立的 $\partial M$ 之定向称为由$M$ 的定向在 $\partial M$ 上诱导的定向 (induced orientation on $\partial M$). 我们现在来具体算出这个诱导定向.

我们仍用上面的记号，并设在 $M$ 中的定向 (其实就是在坐标邻域 $\varphi_\alpha(U)\subset\mathbf{R}^n$ 中的定向) 即

为$(x^1,\cdots,x^n)$. 我们仍然在单形$\sigma$上考虑$\omega$. $\sigma_\alpha$是不必考虑的,因为在$\partial\sigma_\alpha$上$\omega_\alpha=0$,因此我们只来看$U_\beta$和$\omega_\beta$. supp $\omega_\beta$中有一部分$V_\beta$位于$\partial\sigma_\beta$的$\{x^n=0\}$上,除此以外,在$\partial\sigma_\beta$上$\omega_\beta=0$,凡此种种均不必重复,而且下面我们都把下标$\beta$略去. 我们可以假设$\omega=a(x)\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^{n-1}$. 因为$\omega$中其它各项均含$\mathrm{d}x^n$,而在$\partial\sigma$之位于$\{x^n=0\}$上的一部分中自然有$\mathrm{d}x^n=0$,从而$\omega=0$. 在$\partial\sigma$的其它部分上因为已设系数(例如$a(x)$)是有紧支集的从而为0,所以$\omega$中其它各项在$\partial\sigma$上全为0. 这样我们只需考虑

$$\int_{V_\beta}a(x)\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^{n-1}\Big|_{x^n=0}.$$

把它化成$\mathbf{R}^{n-1}$上的黎曼积分成为

$$\int_{\partial\sigma}\omega=\int_{V_\beta}a(x)\Big|_{x^n=0}\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^{n-1}=\varepsilon\int_{V_\beta}a(x)\Big|_{x^n=0}\mathrm{d}x^1\cdots\mathrm{d}x^{n-1},$$

这里$\varepsilon=\pm1$待定. 出现一个$\varepsilon$是因为我们还不知道$V_\beta$上的定向应如何决定才能使斯托克斯定理成立. 于是我们再看$\int_\sigma\mathrm{d}\omega=\int_\sigma\frac{\partial a(x)}{\partial x^n}\mathrm{d}x^n\wedge\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^{n-1}=(-1)^{n-1}\int_\sigma\frac{\partial a(x)}{\partial x^n}\mathrm{d}x^1\cdots\mathrm{d}x^n$. 这里不出现$\varepsilon$是因为我们已规定了$\sigma$中的定向是标准定向$(x^1,\cdots,x^n)$. 于是由$\int_\sigma\mathrm{d}\omega=\int_{\partial\sigma}\omega$应有

$$\varepsilon\int_{V_\beta}a(x)\Big|_{x^n=0}\mathrm{d}x^1\cdots\mathrm{d}x^{n-1}=(-1)^{n-1}\int_\sigma\frac{\partial a(x)}{\partial x^n}\mathrm{d}x^1\cdots\mathrm{d}x^n.$$

因为$a(x)$是有紧支集的,所以不妨认为它定义在整个$H^n$上,而在supp $a(x)$以外恒为0,于是上式成为

$$\varepsilon\int_{\mathbf{R}^{n-1}}a(x)\Big|_{x^n=0}\mathrm{d}x^1\cdots\mathrm{d}x^{n-1}=(-1)^{n-1}\int_{\mathbf{R}^n}\frac{\partial a(x)}{\partial x^n}\mathrm{d}x^1\cdots\mathrm{d}x^n.$$

但这是一个通常的黎曼积分等式,对其右方应用逐次积分,有

$$\int_{\mathbf{R}^n}\frac{\partial a(x)}{\partial x^n}\mathrm{d}x^1\cdots\mathrm{d}x^n=\int_{\mathbf{R}^{n-1}}\mathrm{d}x^1\cdots\mathrm{d}x^{n-1}\int_0^{+\infty}\frac{\partial a(x)}{\partial x^n}\mathrm{d}x^n=-\int a(x)\Big|_{x^n=0}\mathrm{d}x^1\cdots\mathrm{d}x^{n-1}.$$

这里我们利用了$a$具有紧支集,因此当$x^n=+\infty$时$a(x)=0$. 代入上式即得

$$\varepsilon=(-1)^n. \tag{28}$$

这就告诉我们,$\partial M$上的诱导定向应使$\varepsilon=(-1)^n$($\partial M$上的自然定向是$(x^1,\cdots,x^{n-1})$). 这个结论与通常微积分教材上的是一致的,因为通常的微积分教材在这里都用了法向量. 在$\mathbf{R}^n$中有了欧氏结构后(我们的积分理论正建筑在这一点上),切向量$\frac{\partial}{\partial x}$与余切向量$\mathrm{d}x$是一回事(§1),所以$\partial M$即$x^n=0$之内法线向量是$\mathrm{d}x^n$,外法线向量是$-\mathrm{d}x^n$. 再注意到,$\mathbf{R}^n$中的自然定向是$\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n$,而

$$[\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^{n-1}]\wedge\mathrm{d}x^n=(-1)^{n-1}\mathrm{d}x^n\wedge[\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^{n-1}],$$

所以

$$\begin{aligned}[\mathbf{R}^n\text{ 的自然定向}]&=(-1)^{n-1}[\text{内法线}]\wedge[\partial M\text{ 的自然定向}]\\&=-[\text{内法线}]\wedge[\partial M\text{ 的诱导定向}]=[\text{外法线}]\wedge[\partial M\text{ 上的诱导定向}].\end{aligned} \tag{29}$$

就是说，若要求 $\partial M$ 上的诱导定向，先把内法线向量放在待给出 $M$ 的定向的诸向量之末，并且调节其次序使之与 $\mathbf{R}^n$ 的自然一致；再把内法线移到最前面，余下的反号即成 $\partial M$ 上的诱导定向.

看一下 $n=2$ 和 $n=3$ 的情况.

$n=2$ 时，应有 $-[$内法线$]\wedge[\partial M$ 的诱导定向$]=\mathbf{R}^2$ 的右手系. 所以只有 $\partial M$ 的诱导定向为逆时针方向才行. 从这个方向看内法线总是看到它在左侧. 所以微积分教材上都规定：沿曲线正向行走时，区域内域在左侧. $n=3$ 情况有些不同. 图 7－5－7 $n=3$ 的左侧，先是使 $[\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{n}_i]$成右手螺旋即 $\mathbf{R}^3$ 的自然定向. 所以若把 $\boldsymbol{n}_i$ 当作第一个坐标，则诱导定向必为 $[\boldsymbol{t}_2\wedge\boldsymbol{t}_1]$，这样 $\boldsymbol{n}_i\wedge[\boldsymbol{t}_2\wedge\boldsymbol{t}_1]$才与 $\mathbf{R}^3$ 的自然定向相反成左手螺旋系. 但若把内法线 $\boldsymbol{n}_i$ 变成外法线 $\boldsymbol{n}_e$，即从 $M$ 外侧看 $\partial M$，则从 $\boldsymbol{t}_2$ 转到 $\boldsymbol{t}_1$ 成了右手系. 所以在一般的微积分教科书中讲到三维的斯托克斯定理或三维的高斯定理时，常说曲面的定向规定如下：如果“头朝着外法线方向站立”，则正向是右手螺旋系. 这些规定都是通过具体的计算得出来的，而没有说明二维情况与三维情况正向的规定相互有何关系. 至于在更高维的情况，这些非数学的语言如“头朝外法线方向站立”变得毫无意义了，而如何规定 $\partial M$ 上的诱导定向就没有办法讲了. 现在我们的讲法实际上是先证斯托克斯定理，再把能保证此定理能成立的 $\partial M$ 上之定向称为“诱导定向”. 至此为止，和一般的微积分教本讲法一样. 然后，利用外微分形式的知识得到 (29)，即可适用于一般情况.

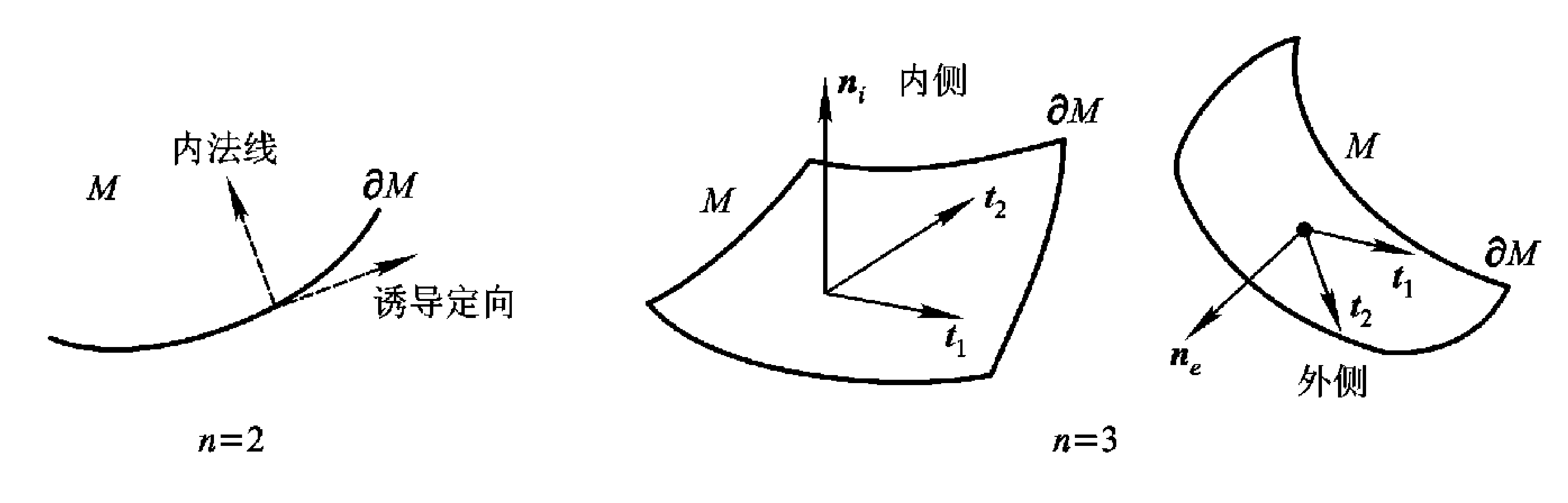



图 7－5－7

以上讲的还不是斯托克斯定理最一般的情况，因为我们考虑的带边的流形至多也只是其一部分区图同胚于半空间 $H^n$ 的情况. 所以它的边界可是说都是非常光滑的. 而一个单形只有其低一维的面才适合这个情况. 单形可以有棱、有角、还有顶点，而对于单形以及由单形构成的链，都可以应用斯托克斯定理. 把这些情况都转移到一般的微分流形，就要求把单形光滑地映射到微分流形上去，这里就用得着微分流形的奇异同调理论，我们也只有在此止步了.

**4. 黎曼流形上的积分**　前面已经说过，要建立积分理论就必须积分区域上能够定义测度. 一般的微分流形虽然是十分光滑的，有很好的拓扑性质，却没有度量性质，因此没有测度. 所以我们甚至无法讲 $\mathbf{R}^n$ 上微分形式的积分，而必须“假想”$\mathbf{R}^n$ 上有了度量 —— 以上我们都是“假想”它有欧几里得度量，然后才“定义”微分形式在 $\mathbf{R}^n$ 上的积分. 如果“假想”其上有别的度量，则积分的“定义”也要修改. 这样，“假想”才有充分理由. 真正的情况是，微分流形上不但可以有度量，而且有很多很多的度量，既然任设一个度量都可以定义微分形式的一种积分，则建立积分理论本来应是毫无问题的事了. 所以下面我们将先证明微分流形 $M$ 上确有度量. 可是，我们在上面还一再提到，微分流形上的积分理论必须是与坐标无关的. 所以，我们上面不是讲函数的积分，而是讲微

分形式的积分，因为微分形式在坐标变换下会自动地提供一个雅可比行列式. 而且当此微分流形为可定向时，可以使这些雅可比行列式均为正，因此会自动出现 $|J|$，这正是在积分中作变量变换之所必须. 现在，如果能找到一个与坐标无关的测度，自然也就可以直接讨论函数 —— 而不是微分形式 —— 的积分了.

现在我们来证明微分流形上确有黎曼度量存在.

**定理 6** 任一微分流形 $M$ 上均有黎曼度量存在.

**证** 在微分流形的定义中即已包含了要求它是豪斯多夫空间以及适合第二可数性公理. 主要的原因之一是为了构造单位分解. 现在设 $\{U_\alpha\}$ 是 $M$ 上一个坐标邻域组成的图册，而 $U_\alpha$ 中的局部坐标为 $(x_\alpha)=(x_\alpha^1,\cdots,x_\alpha^n)$，于是可以作从属于 $\{U_\alpha\}$ 的单位分解 $\{\varphi_\alpha\}$，见(20)式，不过我们现在将 $M$ 上的点 $P$ 标出，而得

$$0\leqslant\varphi_\alpha(P)\leqslant 1,\qquad \sum_\alpha\varphi_\alpha(P)=1.$$

设 $M$ 在 $P$ 点有两个切向量 $v=(v_{\alpha_1}^1,\cdots,v_\alpha^n)$，$w=(w_\alpha^1,\cdots,w_\alpha^n)$，所谓度量就是通过给定的正定矩阵 $(g_{ij}^{(\alpha)}(P))$ 定义 $v$ 与 $w$ 之一个正定的内积：

$$\langle v,w\rangle_\alpha=g_{ij}^{(\alpha)}(P)v_\alpha^i w_\alpha^j.$$

现在用单位分解把各个 $U_\alpha$ 上的内积“拼合”起来，就可以定义切丛 $TM$ 上的内积

$$\langle v,w\rangle(P)=\sum_\alpha\varphi_\alpha(P)g_{ij}^{(\alpha)}(P)v_\alpha^i w_\alpha^j. \tag{30}$$

很容易看出(30)适合内积的要求，特别是正定性，令

$$g_{ij}(P)=\sum_{p\in U_\alpha}\varphi_\alpha(P)g_{ij}^{(\alpha)}(P),$$

即得所求的黎曼度量. 证毕

如果我们令 $g_{ij}^{(\alpha)}(P)=\delta_{ij}$，即在 $P$ 点给出的是欧几里得度量，则黎曼度量其实就是由欧几里得度量拼合起来的. 但是，在 $T_PM$ 上可以有许多不同的内积，对应于各不相同的正交矩阵，所以 $M$ 上也就有无限多种不同的黎曼度量.

现在回到在 $M$ 上如何定义积分的问题. 上述度量是对于作为线性空间的内积来定义的. 即令在 $M$ 上给定了一个度量，还会因为局部坐标的选择而得到不同的表示. 在通常的微积分教本中，我们总是认为欧几里得的度量(即 $g_{ij}=\delta_{ij}$)是最“自然”的，因而讲一维空间上的积分，我们总是用 $\mathrm{d}x$ 作测度. 二重积分则一定是对 $\mathrm{d}x\mathrm{d}y$ 而言. 尽管许多书上一再强调不要把它看成互相正交的 $\mathrm{d}x$ 和 $\mathrm{d}y$ 相乘，而要看成某一个整体 $\mathrm{d}\sigma$，特别讲到某一个区域的可测性 —— 若尔当可测和勒贝格可测 —— 就必须这样. 只在讲到富比尼定理时，才说 $\mathrm{d}\sigma$ 是互相正交的两个轴上的测度 $\mathrm{d}x$ 与 $\mathrm{d}y$ 之积. 但是对于我们，把 $\mathrm{d}x\mathrm{d}y$ 与一个正交的坐标系联系在一起，这个阴影总是挥之不去的. 这是有深刻理由的. 在 §3 定义 6 中，正是借助于 $\mathbf{R}^n$ 中某个度量 $(g_{ij})$ 下的 o. n. 系，来对任意具有相同定向的基底 $\{e_1,\cdots,e_n\}$ 来定义 $\mathbf{R}^n$ 中的体积元 $\sqrt{|g|}e^1\wedge\cdots\wedge e^n$. 对于黎曼流形. 首先是采用黎曼度量 $(g_{ij})$. 因为它是正定的，所以 $|g|=|\det(g_{ij})|=\det(g_{ij})=g$，而且我们就以切空间上的外微分形式来作为 $M$ 上的体积形式.

**定义 5** 设 $M$ 为一可定向的 $n$ 维黎曼流形，$(g_{ij})$ 为其度量，则对任一与此定向相协调的局部坐标系，

$$\eta = \sqrt{g}\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n, \quad g = \det(g_{ij}), \tag{31}$$

称为其典则的(或自然的)体积形式(或体积元).

这个定义是合理的,因为仿§3定义6之说明(见(33)式)知道(31)是与局部坐标的选择无关的.

由于黎曼流形上有度量,所以霍奇 $*$ 算子是可定义的,容易看到

$$*(1) = \sqrt{g}\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n. \tag{32}$$

**定义 6** 若 $f$ 是 $M$ 上一个连续的有紧支集的函数(或 $M$ 为紧),则定义 $f$ 在 $M$ 上的积分为

$$\begin{aligned}\int_M f = \int_M f\eta = \int_M *(f) \\ = \int_M f(x)\sqrt{g}\,\mathrm{d}x^1\cdots\mathrm{d}x^n.\end{aligned} \tag{33}$$

注意(33)的最后一项的写法.当积分区域 $M$ 包含了不只一个坐标邻域,则一定要可以用单位分解将 $f$ 写为 $f = \sum_\alpha f\varphi_\alpha = \sum_\alpha f_\alpha$(而且由 $f$ 具有紧支集或者 $M$ 为紧,这个和只是有限和),而(33)的最后一项应理解为 $\sum_\alpha \int_M f_\alpha$,这个技巧我们已经使用了多次,这里就不再详细说明了.

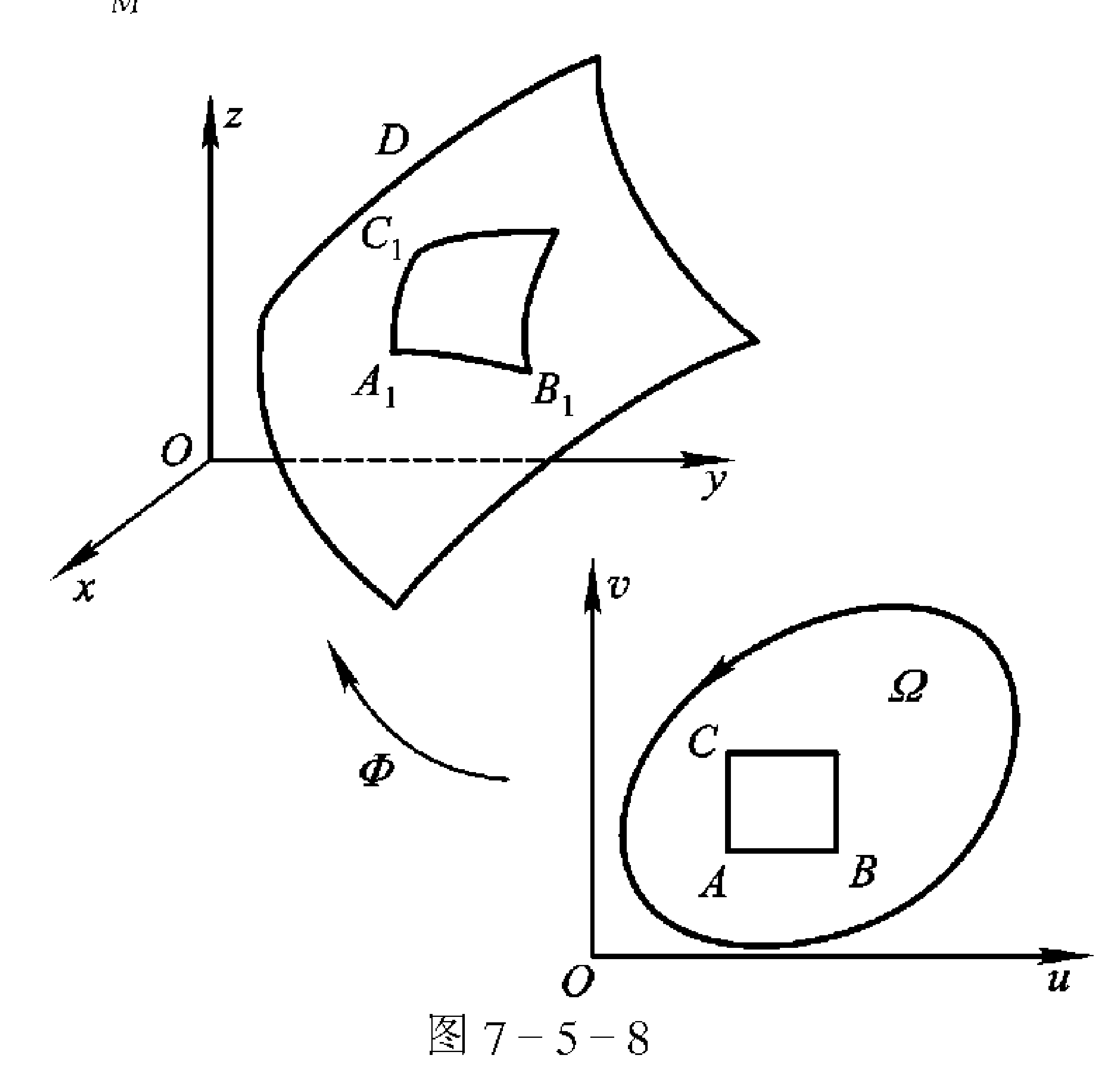


图 7-5-8

这个定义的合理性自然在于 $\eta$ 其实与坐标的选择无关.(33)的第二项 $\int_M f\eta$ 就是一个微分形式 $f\eta$ 在流形 $M$ 上的积分.

我们想回到通常微积分教本中对曲面积分的讲法,这可以帮助我们理解为什么说(31)是体积形式.设有一个曲面 $M$ 在 $\mathbf{R}^3$ 中,我们把 $\mathbf{R}^3$ 看成它的包含空间.在 $M$ 上取一个区域 $D$,并设它对应于参数 $(u,v)$ 空间的区域 $\Omega$(图 7-5-8).于是 $D$ 可以认为是 $\mathbf{R}^2$ 中的区域 $\Omega$ 在映射

$$\Phi: x = x(u,v),\ y = y(u,v),\ z = z(u,v) \tag{34}$$

下的像.设以 $u$ 轴和 $v$ 轴方向上的长为 $\mathrm{d}u, \mathrm{d}v$ 的向量 $\overrightarrow{AB}, \overrightarrow{AC}$ 为棱的矩形被映到 $M$ 上的小区域

$A_1B_1C_1\cdots\cdots$，我们设 $\mathrm{d}u$，$\mathrm{d}v$ 为无穷小量并且来计算 $A_1B_1C_1\cdots\cdots$ 之面积，直到相差一个高阶无穷小量为止. 如果就用 $\mathrm{d}u$，$\mathrm{d}v$ 表示$\overrightarrow{AB}$，$\overrightarrow{AC}$，则 $ABCD$ 之面积是 $*(\mathrm{d}u\wedge\mathrm{d}v)=\mathrm{d}u\mathrm{d}v$. 这里 $*$ 是霍奇算子. 当然用更初等的方法来表示，就是$\overrightarrow{AB}\times\overrightarrow{AC}$. 但是用外微分形式更好，因为我们可以把它们看成切向量(或余切向量，在欧氏空间 $\mathbf{R}^2$ 的笛卡儿坐标系下，二者并无区别)，而当区域 $\Omega$ 被 (34) 映到流形上之区域 $D$ 时，这些切向量将被 $\Phi$ 在 $A$ 点之切映射映到 $M$ 的切空间 $T_AM$ 上，亦即映到 $M$ 的过 $A$ 的切平面上而成为一个平行四边形. 这个切映射在局部坐标 $(u,v;x,y,z)$ 下表为 $\begin{pmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{pmatrix}_A$，所以$\overrightarrow{AB}\,\overrightarrow{AC}$ 之像分别是

$$(x_u\boldsymbol{i}+y_u\boldsymbol{j}+z_u\boldsymbol{k})\mathrm{d}u \text{ 与 } (x_v\boldsymbol{i}+y_v\boldsymbol{j}+z_v\boldsymbol{k})\mathrm{d}v.$$

而 $A_1B_1C_1\cdots\cdots$ 之面积将成为

$$\begin{aligned} &*((x_u\boldsymbol{i}+y_u\boldsymbol{j}+z_u\boldsymbol{k})\mathrm{d}u\wedge(x_v\boldsymbol{i}+y_v\boldsymbol{j}+z_v\boldsymbol{k})\mathrm{d}v) \\ &=[(y_uz_v-y_vz_u)\boldsymbol{i}+(z_ux_v-z_vx_u)\boldsymbol{j}+(x_uy_v-x_vy_u)\boldsymbol{k}]\mathrm{d}u\mathrm{d}v. \end{aligned} \tag{35}$$

它是切平面上的小区域之面积而不是 $M$ 上小区域的面积. 我们说二者相差一个高阶无穷小量，至少应该先已明确了 $M$ 上确有面积存在. 在通常的积分学中，用准确的说法，我们是说这些切平面上小区域之面积和有极限，而面积和与其极限之差为无穷小量，现在我们暂不想去深究. 只想指出 (35) 左方 $*$ 后的外乘积就是面积形式(面积元). 怎样从这里衍生出$\sqrt{g}$ 来呢?下面马上就清楚了.

(35) 是把面积作为一个向量来表出的，但是我们想把它的"数值"，即向量长算出来，经过简单计算，设 $\mathrm{d}u$，$\mathrm{d}v\geqslant 0$，面积数值用 $\mathrm{d}\sigma$ ($\mathrm{d}\sigma\geqslant 0$) 表示，有

$$\begin{aligned} \mathrm{d}\sigma &= \left[\sqrt{(y_uz_v-y_vz_u)^2+(z_ux_v-z_vx_u)^2+(x_uy_v-x_vy_u)^2}\right]\mathrm{d}u\mathrm{d}v \\ &=\sqrt{(x_u^2+y_u^2+z_u^2)(x_v^2+y_v^2+z_v^2)-(x_ux_v+y_uy_v+z_uz_v)^2}\,\mathrm{d}u\mathrm{d}v \\ &=\sqrt{EG-F^2}\,\mathrm{d}u\mathrm{d}v. \end{aligned} \tag{36}$$

$$E=x_u^2+y_u^2+z_u^2=\left|\overrightarrow{A_1B_1}\right|^2/(\mathrm{d}u)^2,\ G=x_v^2+y_v^2+z_v^2=\left|\overrightarrow{A_1C_1}\right|^2/(\mathrm{d}v)^2,$$

$$F=x_ux_v+y_uy_v+z_uz_v=(\overrightarrow{A_1B_1}\cdot\overrightarrow{A_1C_1})/\mathrm{d}u\mathrm{d}v.$$

$E$，$F$，$G$ 的表达式在通常的微积分教本中都有，只是没有指出例如 $E=(\overrightarrow{A_1B_1})^2/(\mathrm{d}u)^2$ 而已.

但是，用向量表示面积终属不便. 为此，我们在 $M$ 上 $A_1$ 点取一个单位法线向量 $\boldsymbol{n}$，并使 $(\overrightarrow{A_1B_1},\overrightarrow{A_1C_1},\boldsymbol{n})$ 成为右手系. (或者说选 $\boldsymbol{n}$ 就是$\overrightarrow{A_1B_1}\times\overrightarrow{A_1C_1}$ 方向的单位向量). 也就是说，如果在 $(u,v)$ 平面上再加上 $w$ 轴使 $(u,v,w)$ 成右手系，则 (34) 保持定向不变. 我们通常的说法是面积元 $\mathrm{d}\sigma$ 是在曲面 $M$ 上 $\boldsymbol{n}$ 所指的一侧. 这样做最大的优点是面积将始终为正. 因为这时以 $(\overrightarrow{A_1B_1},\overrightarrow{A_1C_1},\boldsymbol{n})$ 为棱的平行六面体之体积(它一定是正的，因为三个棱的定向与 $\mathbf{R}^3$ 中右手坐标系定向一致) 之数值与底面积 $A_1B_1C_1\cdots\cdots$ 数值一致. 而体积是用混合积表示的：

$$*(\overrightarrow{A_1B_1}\wedge\overrightarrow{A_1C_1}\wedge\boldsymbol{n})=\boldsymbol{n}\cdot(\overrightarrow{A_1B_1}\times\overrightarrow{A_1C_1})=\left|\overrightarrow{A_1B_1}\times\overrightarrow{A_1C_1}\right|.$$

最后一式的来源是$\overrightarrow{A_1B_1}\times\overrightarrow{A_1C_1}$ 与 $\boldsymbol{n}$ 有相同方向. 按通常的向量公式计算又得 (36) 式. 总之我们得到了一个正的面积的公式，而通常微积分教本中曲面面积公式

$$\iint_{\Omega}\sqrt{EG-F^2}\,\mathrm{d}u\mathrm{d}v$$

及其符号规定均由此而来.

$E,F,G$ 和度量有什么关系呢?在 $\mathbf{R}^3$ 中有欧几里得度量

$$\mathrm{d}s^2=\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2.$$

以(34)代入,即得 $M$ 上的一个度量

$$\begin{aligned}\mathrm{d}s^2&=(x_u\mathrm{d}u+x_v\mathrm{d}v)^2+(y_u\mathrm{d}u+y_v\mathrm{d}v)^2+(z_u\mathrm{d}u+z_v\mathrm{d}v)^2\\&=E\mathrm{d}u^2+F\mathrm{d}u\mathrm{d}v+F\mathrm{d}v\mathrm{d}u+G\mathrm{d}v^2.\end{aligned}$$

可见 $E=g_{11},F=g_{12},G=g_{22}$,而 $\sqrt{EG-F^2}=\sqrt{g}.g=|g|$,因为 $g$ 是正定的.可见 $\sqrt{EG-F^2}\mathrm{d}u\wedge\mathrm{d}v$ 就是 $M$ 上之面积形式(面积元),$\sqrt{EG-F^2}\mathrm{d}u\mathrm{d}v=*(\sqrt{EG-F^2}\mathrm{d}u\wedge\mathrm{d}v)$ 就是面积.

不过我们在 $M$ 上给出的度量是由 $\mathbf{R}^3$ 中的欧几里得度量衍生(诱导)而来的.通常的曲线 $L$ 也是作为一个一维流形,放在包含空间 $\mathbf{R}^3$ 中,由后者的欧几里得度量即可诱导出 $L$ 之通常的弧长公式.如果考虑一般的流形,其度量不一定是由包含空间的欧几里得度量诱导而来,我们就说 $\sqrt{g}\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n$ 是其体积形式(体积元).也应该说 $*(\sqrt{g}\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n)=\sqrt{g}\mathrm{d}x^1\cdots\mathrm{d}x^n$ $(\mathrm{d}x^i\geqslant 0)$ 是体积.

求曲线弧长时,例如求圆周长时,我们既可以用内接正多边形求极限,也可以用外切多边形求极限.第一章介绍穷竭法时就讲到这一点.可是求曲面面积则不一样.很早就有例子表明,一个曲面简单如圆柱面,若用内接多面体求极限甚至会求出无穷大!但是对可微曲面,若用外切多面体求极限就不会有这样的问题.好在我们是用 $(u,v)$ 平面上的 $ABC\cdots\cdots$ 经切映射得到曲面的切平面上的 $A_1B_1C_1\cdots\cdots$,所以不会有这样的困难.

然而,研究黎曼流形上的函数的积分,主要目的却不在于此.现在,我们想要在黎曼流形 $M$ 上建立 $L^2$ 理论.为了建立这样一个理论,本来不必要求 $M$ 是黎曼流形,但是确实需要在 $M$ 上能够定义正的测度.对一般的微分流形,当然也可以仿照定理 6 的作法,先作一个坐标邻域的图册,而在每个坐标邻域 $U_\alpha$ 中利用局部坐标定义一个测度,例如勒贝格测度 $\mathrm{d}\mu=\mathrm{d}x^1\cdots\mathrm{d}x^n$,再用单位分解把它们拼合起来,只不过这样作出来的测度没有对坐标的不变性罢了.而我们作出的黎曼度量以及体积形式也就给了我们一个正的测度 $\mathrm{d}\sigma=*(\sqrt{g}\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n)=\sqrt{g}\mathrm{d}x^1\cdots\mathrm{d}x^n$,所以我们仍设 $M$ 为一黎曼流形,其上有度量 $(g_{ij})$.

为了建立 $L^2$ 理论,我们不必如第四章中那样从可测函数开始,而只要注意到,$L^2$ 理论的基础是定义内积.如果有了两个 $p$ 微分形式 $\omega$ 和 $\sigma$,而可以定义其内积 $(\omega,\sigma)$(请读者特别注意,我们这里没有用 $\langle\omega,\sigma\rangle$ 这样的记号,并不是由于现在的 $\omega$ 与 $\sigma$ 是复的对象,而是因为我们在 §3 定理 8 中已用了 $\langle\omega,\sigma\rangle$ 这个记号而且马上就要再引用它),因此也就可以定义其范数 $\|\omega\|^2=(\omega,\omega)$,本来,我们讲的所有微分形式都是 $C^\infty$ 的,但是有了范数以后,就可以把它们按此范数完备化而得出 $L^2$ 空间.至于 $(\omega,\sigma)$ 的作法,在 §3 中已在一个 $n$ 维线性空间 $\mathbf{R}^n$ 中定义了 $\langle\omega,\sigma\rangle$.现在取此空间为 $T_PM$,则 $\langle\omega,\sigma\rangle$ 之值因 $P$ 而异成了一个函数.因为 $\omega,\sigma$ 是 $C^\infty$ 的外微分形式,所以 $\langle\omega,\sigma\rangle(P)$ 是 $M$ 上的 $C^\infty$ 函数,今设它们都有紧支集(或者设 $M$ 为紧)——这个假设的作用在于回避反常积分,前面已见过多次——我们有

**定义 7** 设 $M$ 为一可定向的黎曼流形. $\omega$ 和 $\sigma$ 为其上的紧支集 $C_0^\infty$ $p$ 微分形式(或设 $M$ 为紧),则可以定义它们的内积为

$$(\omega,\sigma)=\int\langle\omega,\sigma\rangle=\int\langle\omega,\sigma\rangle(x)\sqrt{g}\,\mathrm{d}x^1\cdots\mathrm{d}x^n, \tag{37}$$

然后就可以用完备化定义 $M$ 上的平方可积函数空间 $L^2(M)$.

有了 $L^2(M)$ 的理论,也就是有了 $M$ 上的希尔伯特空间理论. 而且还可以进一步定义 $M$ 上的索波列夫空间,可以定义 $M$ 上的广义函数理论等等. 可是我们现在只想简单地讲一下 Laplace-Baltrami 算子.

上节末我们把 Laplace-Beltrami 算子定义为

$$\Delta=\mathrm{d}\delta+\delta\mathrm{d}:\quad F^p(\mathbf{R}^n)\to F^p(\mathbf{R}^n),$$

这里的 $\delta$ 在上节中说是 d 的"对偶"算子,其定义是 §4 的(37) 式. 其中用到了霍奇 $*$ 算子,其定义及性质见 §3 定理 9 与定理 10,其中用到一个参数 $s$ 是由于定义 $*$ 时要考虑到 $\mathbf{R}^n$ 中的度量. 度量是用一个非蜕化对称矩阵来定义的. 对称矩阵只有实特征根,其中正特征根与负特征根个数都是重要的不变量. $s$ 就是负特征根个数. 现在我们用的黎曼度量相应于正定矩阵 $(g_{ij})$,所以 $s=0$. 这样由 §3 的(39),

$$**=(-1)^{p(n-p)},$$

而有

$$*^{-1}=(-1)^{p(n-p)}*. \tag{38}$$

那么,上节说 $\delta$ 是 d 的对偶是什么意义呢?例如 $n$ 阶矩阵 $A:\mathbf{R}^n\to\mathbf{R}^n$ 的转置矩阵 ${}^tA:\mathbf{R}^n\to\mathbf{R}^n$,可以由

$$\langle Au,v\rangle=\langle u,{}^tAv\rangle$$

来定义. $\langle\cdot,\cdot\rangle$ 是 $\mathbf{R}^n$ 的内积. 现在我们考虑的是 $F^p(\mathbf{R}^n)$,其中的内积由(37) 定义. 我们要证明 d 与 $\delta$ 的关系就如同矩阵 $A$ 与 ${}^tA$ 的关系一样:

$$(\mathrm{d}\omega,\sigma)=(\omega,\delta\sigma). \tag{39}$$

这里 $\omega,\sigma$ 是具有紧支集的 $C^\infty$ 的 $F^{p-1}(M)$ 与 $F^p(M)$ 中之元. 但与 $\mathbf{R}^n$ 的情况不同,那里的 $u,v$ 应是 $\mathbf{R}^n$ 中的任意元,现在一方面例如 d 涉及微分,所以对任意的 $\omega\in L^2(M)$,$\mathrm{d}\omega$ 不一定有意义,再则(39) 很像分部积分法,这一点与第四章讲的广义函数一比较就清楚了. 要作分部积分就会出现积分号外之项,而我们又必须设法消除它们. 为此我们设 $\omega,\sigma$ 均为有紧支集的 $C_0^\infty$ 的外微分形式. 这样我们就绕过了 d 与 $\delta$ 的定义域究竟是什么这个问题. 而回避了这个问题以后,我们就说 $\delta$ 是 d 的形式伴算子(formal adjoint operator),这样我们就有

**定理 7** $\delta$ 是 d 的形式伴算子,即若设 $\omega\in F^{p-1}(M)$,$\sigma\in F^p(M)$ 均为 $C^\infty$ 的而且有紧支集,则有下式成立:

$$(\mathrm{d}\omega,\sigma)=(\omega,\delta\sigma). \tag{39}$$

**证** 由 $*$ 的定义,$*\sigma\in F^{n-p}(M)$,从而 $\omega\wedge*\sigma\in F^{n-1}(M)$,而 $\mathrm{d}(\omega\wedge*\sigma)$ 有意义. 由外微分算子之性质有

$$\begin{aligned}\mathrm{d}(\omega\wedge*\sigma)&=\mathrm{d}\omega\wedge*\sigma+(-1)^{p-1}\omega\wedge\mathrm{d}*\sigma\\&=\mathrm{d}\omega\wedge*\sigma-\omega\wedge*[(-1)^p*^{-1}\mathrm{d}*\sigma]\\&=\mathrm{d}\omega\wedge*\sigma-\omega\wedge*\delta\sigma=\langle\mathrm{d}\omega,\sigma\rangle-\langle\omega,\delta\sigma\rangle.\end{aligned}$$

这里我们用了 §4 中 $\delta$ 的定义(39) 式以及 $**^{-1}=\mathrm{id}$. 双方在 $M$ 上积分并用斯托克斯定理,即知左方积分为 0 而得(39) 式,定理证毕.

这个定理的证明看起来只是形式演算,实际上与用分部积分法是一样的. 说到底,分部积分法也只是一个形式演算.

最后我们再来看 Laplace-Beltrami 算子的具体表示. §4 中当度量矩阵是对角阵时给出了所需结果,现在要看对于一般的黎曼度量 $(g_{ij})$ 又当如何. 我们仍然只看对于 $u\in F^0(M)$,$\Delta u$ 怎样表示. 注意到 $\delta: F^p(M)\to F^{p-1}(M)$,当 $p=0$ 时 $F^0(M)$ 即 $C^\infty$ 函数,而 $F^{-1}(M)$ 只好说是 0 元素的空间,因此 $\delta u=0$. 这样 $\Delta u=\delta\mathrm{d}u$. 我们假设 $u\in F^0(M)$,但设 $\varphi\in C_0^\infty(M)$,这与广义函数论中的试验函数是一致的. 于是

$$\begin{aligned}\int_M \Delta u\cdot\varphi\sqrt{g}\,\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^n &= (\delta\mathrm{d}u,\varphi)=(\mathrm{d}u,\mathrm{d}\varphi)\\ &=\int_M\langle\mathrm{d}u,\mathrm{d}\varphi\rangle\sqrt{g}\,\mathrm{d}x^1\cdots\mathrm{d}x^n\\ &=\int_M g^{ij}\frac{\partial u}{\partial x^i}\frac{\partial\varphi}{\partial x^j}\sqrt{g}\,\mathrm{d}x^1\cdots\mathrm{d}x^n\\ &=-\int_M\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^j}\left(\sqrt{g}g^{ij}\frac{\partial u}{\partial x^i}\right)\varphi\sqrt{g}\,\mathrm{d}x^1\cdots\mathrm{d}x^n.\end{aligned}$$

所以有

$$\Delta u=-\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^j}\left(\sqrt{g}g^{ij}\frac{\partial u}{\partial x^i}\right).\tag{40}$$

我们前面已证明了

$$\Delta=-\operatorname{div}\mathbf{grad},$$

与(40) 比较,我们应定义 $u$ 之梯度是一个向量场:

$$\mathbf{grad}\ u=g^{ij}\frac{\partial u}{\partial x^i}\frac{\partial}{\partial x^j}.\tag{41}$$

这是由于这样定义 $\mathbf{grad}\ u$ 后,对任意向量场 $X$ 我们有

$$\langle\mathbf{grad}\ u,X\rangle=X(u)=\mathrm{d}u(X).\tag{42}$$

而且对于向量场 $X=X^i\dfrac{\partial}{\partial x^i}$,应定义

$$\operatorname{div}X=\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^j}(\sqrt{g}X^j)=\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^j}\left(\sqrt{g}g^{ij}\left\langle X,\frac{\partial}{\partial x^i}\right\rangle\right).\tag{43}$$

这样一来

$$\Delta u=-\operatorname{div}\mathbf{grad}\ u$$

仍成立.

但是对高次的微分形式 $u$,$\Delta u$ 的计算是比较复杂的. 它涉及许多较深入的微分几何问题.

# §6 结束语——麦克斯韦方程组简介

本书即将结束了.回顾一下,我们在很简略地介绍了一点希腊数学中有关极限理论的研究以后,很快地就转入了微分学和积分学.我们介绍了牛顿如何从开普勒三定律得出引力定律,解决了当时(17世纪中叶)科学中最重要的问题:太阳系的运行规律.而微积分又如何由此形成一门系统的科学.我们现在通用的微积分教材中常讲的一些静力学问题时代更早,都是17世纪前科学中的重要问题.几何光学虽然稍晚一些也差不太多.应用范围虽然很大,出现的数学问题虽然形形色色,解决这些问题的思想都是一样的:都涉及如何处理无穷小量:对于微分学,就是允许略去高阶无穷小量.对于积分学实际上是"无穷多个无穷小"如何求和的问题.虽然从希腊人起就知道存在这些问题,但是过多的思辨、抽象的议论(在当时条件也只能如此)使得对这些问题的解决完全不能适应数学发展的需要.一直到19世纪中下叶,人们才摆脱了那些抽象的哲学思辨,以及缺少分析的直觉.现在通用的微积分教本的讲法正是反映了19世纪下半叶我们对无穷小、极限等概念的理解:既摆脱了抽象的哲学思辨,也摆脱了缺少分析——主要指逻辑上的分析——的直觉.我们掌握了它,也就可以有把握地前进了.这一段历史表明了数学的发展与牛顿所实现的科学上的第一次大综合的关系何等密切.由于数学发展到如此成熟的地步,我们今天甚至可以使用一些很直观的语言,而不必担心会产生什么问题.本书里对微分学和积分学的讲法也无非是从今天的角度再次解释这些基本概念的现代的理解和表述.说到底还是牛顿的伟大成就.由于是牛顿的成就的反映,所以还有一个非常重要的特点是:解决各种物理问题,都是在牛顿的框架思路之下:问题本身都很明确,解决问题所用到的数学理论也很明确,没有太大的变化.当然有些问题难度很大.现在的问题是,自微积分学系统地形成了以后,有没有更重要的新发展?我们在本书中实际上介绍了一些重要发展,例如变分法、勒贝格积分及其在研究随机现象上的应用,还有调和分析.它们都有自己的非常广阔的前途,都已经超越了牛顿的框架.但下面我们还要介绍一下麦克斯韦(以下都简记为M)关于电磁场的理论.一方面因为它可以说是物理科学自牛顿以后第二次伟大的综合,而且直接为20世纪物理学的大革命——相对论的出现——开辟了道路:M理论本质上是相对论性的.更重要的是它所需要的数学与牛顿时代大不相同了.19世纪中叶,自从高斯、黎曼以后,空间的本性问题以一种远远超越哲学思辨的形态提到人们面前.空间既然已不一定限于三维平直的空间,则人们很自然地先是考查高维但是平直的空间——线性空间.于是我们有了线性代数,它的作用远不止我们在微分学一章中指出的线性化,即作为一般非线性问题的近似.它是一个伟大框架前提的一部分.人们一直试图寻找新的刻画大自然的数学工具,例如哈密顿(W.R.Hamilton,一位非常重要的数学家和物理学家)研究四元数就是这个企图的表现.那时,四元数十分为人们看好.例如M本人还有凯尔文勋爵(Lord Kelvin,真名是William Thomson)都想用四元数来作为研究电磁现象的基本工具.而现在知道四元数的人已经不多了.线性代数的研究看来一直是探讨空间本性的大潮流的一个部分,不过因为它是考察一些最基本的结构,而当时这批数学家所关心的应用问题我们又不甚熟悉,所以现在人们都感到它太抽象,似乎与牛顿大相径庭了.其实并不全然如此.这个潮流中有一位很深刻的数学家格拉斯曼.他的重要性似乎今天大多数读者估计不足.到20世纪初,开拓了比较系统的关于弯曲空间的研究,有

了微分流形理论，有了黎曼几何等等，终于诞生了爱因斯坦和他的相对论. 我们不妨说，高斯黎曼播种，爱因斯坦收获. 在一个多世纪后的今天看来，他们播种的是一种数学框架，里面包含了许多新概念，新方法. 电磁现象要放在这样一个框架中才能伸展自如. 牛顿的摇篮放不下 M 这个巨人，更不说爱因斯坦了. 这一个框架的内容之丰富决非本书所能容纳. 以后还发展了如规范场理论等等也都是内容更丰富、更深刻也就需要新的数学的框架. 下面我们介绍 M 的理论是一个绝大多数学生都能懂而又十分重要的课题，可以就此看一下这个新的数学框架的威力.

M 关于电磁现象的研究是继承了法拉第(M. Faraday，1791—1867) 的工作. 法拉第对电磁现象的理解与牛顿对引力的理解不同. 牛顿的理论承认超距作用(尽管他自己也为此苦恼)：例如两个天体的吸引是瞬时的，引力的传播不需要时间，也不需要某种介质的媒介. 法拉第则不同，他认为电和磁的影响是通过"场" 来实现的. 电场由电力线组成. 磁场由磁力线组成、电力线与磁力线并不像某些中学教材里说的，只是方便的虚构，而是实实在在的，甚至有弹性，可以互相吸引或排斥. M 的工作正是从深刻研究法拉第的著作《电的实验研究》开始，而且十分明确地把法拉第的方法与按超距作用理论进行探讨的"数学家" 们对立起来. 他说："在他的心目中，法拉第看到一些力线穿过全部的空间，而数学家们则只在空间中看到一些超距吸引着的力心；法拉第看到一种媒质，而他们则除距离以外毫无所见；法拉第向在媒质中进行着的真实作用中寻找现象的依据，而他们则满足于在对媒质发生超距作用的一种本领中找到了这种依据". (见 M 的名著：A Treatise on Electricity and Magnetism，1873，中译本：《电磁通论》(上)，xii 页，武汉出版社 1992，以下引用此书时均依据这个译本). M 的工作就是把法拉第的思想数学化. 他的成就得到了法拉第高度评价.《电磁通论》就是总结这些工作的一部划时代的名著. 于是 M 认为电磁现象应该用电磁场来描述. 电场和磁场分别用向量 $\boldsymbol{E}(x, y, z, t)$ 和 $\boldsymbol{B}(x, y, z, t)$ 来表示. 它们在某种介质中传播，这种介质就称为以太(ether). 以太被想像成为一种穿透整个宇宙的弹性体. 它不可见，不可损坏，然而却有一种奇怪的特性，即只允许横波通过. 就本书涉及的大人物而言，纳维埃、柯西、格林，斯托克斯都赞成以太学说，而 M 对电磁场的传播最早给出的一个模型纯粹是横波在弹性介质中传播的模型，并且计算出传播速度就是光速. 光速是电磁理论中极重要的常量，本章恒用 $c$ 表示. 就此，M 写道："我们很难避开一个推论，即光是造成电磁现象的那一种介质中的横振动"(1862).

前面我们已提到，M 是最早主张用向量而不是四元数来刻画电磁现象的物理学家之一. 他指出，由一个向量场 $\boldsymbol{A}(x, y, z, t)$ 一定可以产生两个量：一是它沿某一曲面一定的法线方向 $\boldsymbol{n}$ 穿过此曲面 S 的通量(flux)：

$$\iint_S \boldsymbol{A}\cdot\boldsymbol{n}\mathrm{d}\sigma = \iint_S A_1\mathrm{d}y\mathrm{d}z + A_2\mathrm{d}z\mathrm{d}x + A_3\mathrm{d}x\mathrm{d}y;$$

另一个是它沿某一有一定定向的闭曲线 $L$ 的环流(circulation)：

$$\int_L \boldsymbol{A}\cdot\mathbf{d}\boldsymbol{s} = \int_L A_1\mathrm{d}x + A_2\mathrm{d}y + A_3\mathrm{d}z.$$

M 十分注意这两个概念在电磁现象的研究中的作用. 因为他说"在电的流动事例中，我们根本不知道有关导体中的电密度或电速度的任何东西，我们只知道按照流体理论将对应于密度和速度之乘积的那个值. 因此，在所有的这种事例中，我们必须应用测量通过面积之通量的那种方法."(同上，13 页) 这就是说，我们需要知道的是通过某一个小面积 $\sigma$ 的电流(这是一个向

量) $\boldsymbol{j}\mathrm{d}\sigma$. $\boldsymbol{j}$ 称为电流密度向量，这是今天我们使用的语言. 因此，电通过曲面 $S$ 的电流即 $\boldsymbol{j}$ 的通量

$$\iint_S \boldsymbol{j}\,.\,\boldsymbol{n}\mathrm{d}\sigma.$$

至于电荷的密度 $\rho$，则没有任何困难.

考虑 $\boldsymbol{j}$ 的通量即得到一个基本的事实——电荷守恒. 设曲面 $S$ 包围了区域 $\Omega$，而有电流流入，电流密度向量为 $\boldsymbol{j}$ (图 7－6－1)，于是在时间 $\mathrm{d}t$ 内由 $\Omega$ 外经过 $S$ 流入 $\Omega$ 的电量 (注意，M 一直是以流体作为电的模型的) 是

$$-\mathrm{d}t\iint_S \boldsymbol{j}\cdot\boldsymbol{n}\mathrm{d}\sigma.$$

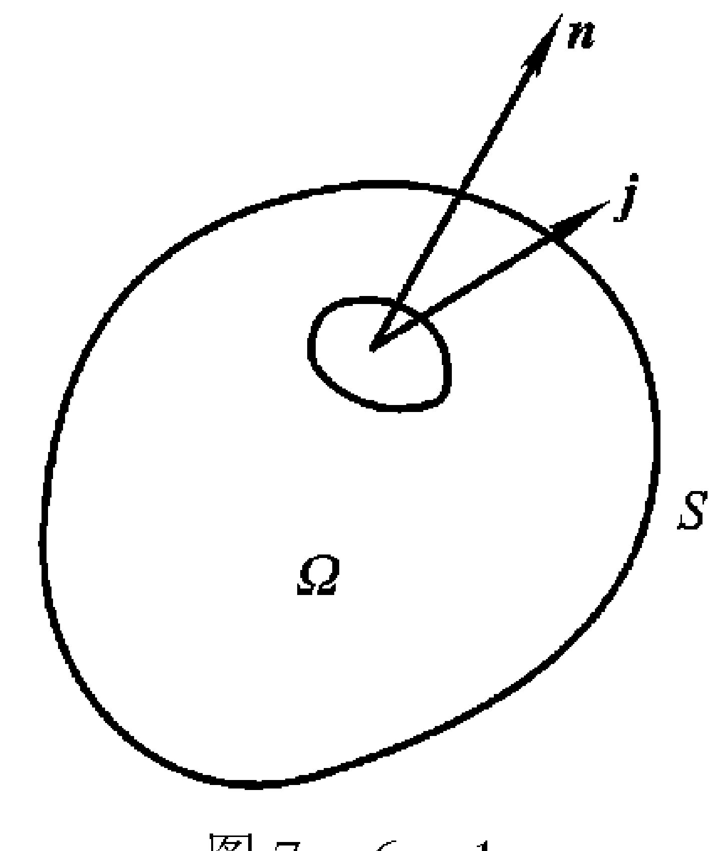


图 7－6－1

它造成 $\Omega$ 内电荷的积累，即电荷应增加

$$\mathrm{d}t\iiint_\Omega \frac{\partial\rho}{\partial t}\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

这两个量应相等. 由于此关系应在任意区域 $\Omega$ 上都成立，故由斯托克斯定理 (在通常的微积分教本中应用于这一情况的斯托克斯定理称为高斯定理) 有

$$\mathrm{d}t\iiint_\Omega \left[\frac{\partial\rho}{\lambda t}+\mathrm{div}\,\boldsymbol{j}\right]\mathrm{d}x\mathrm{d}y\mathrm{d}z = 0.$$

因此得到连续性方程

$$\frac{\partial\rho}{\partial t}+\mathrm{div}\,\boldsymbol{j} = 0. \tag{1}$$

这是电磁现象第一个基本的事实——电荷守恒.

除此以外，M 总结了当时关于电磁现象的全部研究成果，提出了著名的 M 方程组. 这里应该指出 M 与牛顿的不同. 牛顿在《原理》一书中是把他的三定律当作欧几里得的《几何原本》中的公理对待，而由此推导出其它一切结论. M 则不同，他是把以前的全部成果特别是实验成果整理归结为四个方程，后来人们才称为 M 方程组. 而且如电荷守恒这样最基本的事实也没有列入 M 方程组中.

首先看电场 $\boldsymbol{E}$ 与磁场 $\boldsymbol{B}$ 的通量. 关于电场通量，高斯早就证明了，穿过封闭曲面 $S$ 的 $\boldsymbol{E}$ 之通量等于 $\Omega$ 中的电荷. 于是应用推导连续性方程的方法即得

$$\mathrm{div}\,\boldsymbol{E} = 4\pi\rho. \tag{Ⅰ}$$

注意，M 方程组的写法各书常有不同，这里是讨论真空中的场，并采用了高斯——MKS 单位制[①]. 其实方程 (Ⅰ) 高斯早已得出，并不是 M 的方程.

用同样方法处理磁场 $\boldsymbol{B}$，由于磁单极是不存在的：有磁南极则必有大小相同的磁北极在一起，因此总磁荷必为 0. 这样相应于 (1) 必有

$$\mathrm{div}\,\boldsymbol{B} = 0. \tag{Ⅱ}$$

① 这里提到单位问题. 我们念数学的人有一个大弱点，即对单位不注意. 单位问题首先是一个“应用问题”. 因此，任何一个国家都有相应的规定. 我国国家技术监督局就不把 MKS 制作为标准单位制，而是作为常用单位制，因为标准单位一般用于产业中. 现在的物理教材则常采用高斯－MKS 制. 但单位问题还有物理意义. 从库伦定律和安培定律采用的单位的分析中就可以导出光速来，这一点很值得注意.

M 方程组的另外两个方程是关于 $\boldsymbol{B}$ 和 $\boldsymbol{E}$ 的环流的.为此设 $L$ 是一个封闭曲线,它是一"片"曲面 $S$ 的边界:$L = \partial S$,而且 $L$ 上的正向与 $S$ 上的某一指定的法线方向 $\boldsymbol{n}$ 成为右手坐标系.于是斯托克斯定理给出

$$\int_L \boldsymbol{A} \cdot \mathbf{d}\boldsymbol{s} = \iint_S (\mathbf{curl}\ \boldsymbol{A})_n \mathrm{d}\sigma.$$

现在把它用于 $\boldsymbol{E}$.法拉第的电磁感应定律告诉我们:磁场对时间的变化会引起线圈中出现电流.用数学方式来表述,即($\boldsymbol{E}$ 绕 $\partial S$ 之环流)$= -\dfrac{1}{c}\left(\dfrac{\partial \boldsymbol{B}}{\partial t}\right.$ 穿过 $S$ 之通量$\left.\right)$

$$\int_{\partial S} \boldsymbol{E} \cdot \mathbf{d}\boldsymbol{s} = -\frac{1}{c}\iint_S \left(\frac{\partial \boldsymbol{B}}{\partial t}\right)_n \mathrm{d}\sigma.$$

应用斯托克斯定理有

$$\iint_S (\mathbf{curl}\ \boldsymbol{E})_n \mathrm{d}\sigma = -\frac{1}{c}\iint_S \left(\frac{\partial \boldsymbol{B}}{\partial t}\right)_n \mathrm{d}\sigma.$$

由于 $S$ 的任意性(所以其法线方向 $\boldsymbol{n}$ 也是任意的),故有

$$\mathbf{curl}\ \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t}. \tag{Ⅲ}$$

最后一个也是最值得注意的是表明电流可以产生磁场,这要从 1819—1820 年奥斯特(H. C. Oersted)的实验讲起.当时他从一个模糊的思想:磁可能是电的一种"潜在形式"出发,发现了通过电流的导线——这种电流不妨称为传导电流——附近会观察到磁针的偏转.电流的磁效应的发现引起了科学界极大的重视.紧接着有安培、毕奥-萨伐尔的定律等等,但都是表明了电流的磁效应.这样电与磁明显地处于一种对称的、互相产生的关系中.奥斯特的实验以后的 12 年即 1831,法拉第就发表了他的电磁感应定律.M 后来把它表示成(Ⅲ),而按奥斯特的实验的数据结果来分析,应该得到

$$\boldsymbol{B} \text{ 绕 } \partial S \text{ 的环流} = \frac{4\pi}{c}(\boldsymbol{j} \text{ 通过 } S \text{ 的通量}).$$

这样一来,仿照上面得出法拉第感应定律的方法应有

$$\mathbf{curl}\ \boldsymbol{B} = \frac{4\pi}{c}\boldsymbol{j}. \tag{2}$$

很遗憾,这是错误的.科学的发展告诉我们,千万不要无视"错误"——只要它是一个严肃的错误.正是从这个错误中 M 完成了他的伟大功绩.我们看到:(2)与(Ⅲ)明显地不对称.这件事引起了法拉第的注意:如(Ⅲ)那样,是$\dfrac{\partial \boldsymbol{B}}{\partial t}$决定了 $\boldsymbol{E}$,而恒定的磁场并不产生电场,所以(2)中缺少了$\dfrac{\partial \boldsymbol{E}}{\partial t}$这样的项.M 是把电场与一种连续介质相比拟的.因此,他不仅要考虑真空中的电场,而且考虑电介质中的电场,并且认为电动势存在时,介质如同弹性体一样受到应力(他称为胁强),产生一种称为电位移的东西.M 说:"电位移的变化显然就构成电流.然而这种电流只有在电位移变化的过程中才能存在.…… 不能像导体中的电流那样不受限制地沿着相同的方向继续流动"(同上,72 页).这就是说,产生磁场的电流有两个部分,一是"像导体中的电流"即传导电流 $\boldsymbol{j}$,另一部分则是电位移的变化,称为位移电流.按这样的思考,M 指出,(2)其实是漏掉了位移电流,而正确的方程应该是

$$\textbf{curl}\ \boldsymbol{B} = \frac{1}{c}\frac{\partial \boldsymbol{E}}{\partial t} + \frac{4\pi}{c}\boldsymbol{j}. \tag{Ⅳ}$$

这真是神来之笔!尽管 M 是从一个不合适的连续介质模型出发,可是得到了一个十分重要的结论,它有十分重要的推论,而且得到了实验的证明. 总之,M 方程组现在是完备了,它成为电动力学(电磁场随着时间变化,所以称为动力学)的基础.

由于位移电流概念的极大的重要性,我们再从数学角度来考察它,看一下为什么没有位移电流会产生矛盾. 设有电路如图 7 - 6 - 2,导线上有传导电流 $\boldsymbol{j}$,而导线连接着电容器的两极. 电容器的两极间不是导体,而可能是某种媒质,其中没有传导电流. 我们按(2)那样来分析这个电路,作一条绕着导线的曲线 $L$,并以 $L$ 为边缘作一个曲面 $S_1$,它与导线相交. 在 $S_1$ 上指定某一侧的法线为 $\boldsymbol{n}$. 于是 $L = \partial S$,而且 $L$ 上有诱导定向,如图上实线箭头所示. 诱导定向是由 $\mathbf{R}^3$ 中指定了一个定向以后才有的. 我们指定的是 $\mathbf{R}^3$ 中用右手系. 但实际上即使指定为左手系也无关,最后会得出同样的矛盾. 研究 $\boldsymbol{B}$ 在 $L$ 上的环流,于是按(2)式,知有传导电流 $\boldsymbol{j}$ 穿过 $S_1$ 其通量为 $\iint_{S_1} \boldsymbol{j}\cdot\boldsymbol{n}\mathrm{d}\sigma$(这里有一点小问题,即 $\boldsymbol{j}$ 只在导线上有,而上面的积分应按 $\delta$ 函数去理解,我们不去管这件事,因为它确实是很容易解决的),总之该通量不为 0.

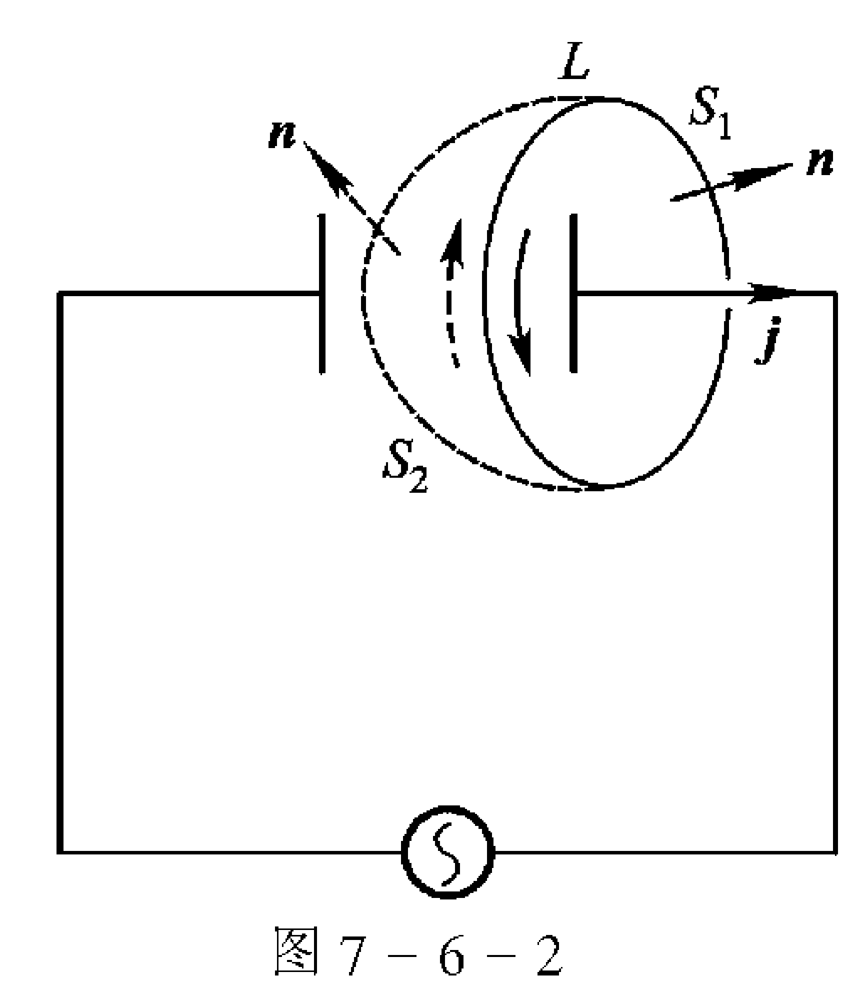



图 7 - 6 - 2

另一方面,我们可以再穿过电容器的两壁,作一个曲面 $S_2$,仍以 $L$ 为边缘:$L = \partial S_2$,但过 $S_2$ 并没有电流,因此按(2),应有

$$(\boldsymbol{B}\ \text{沿}\ L\ \text{的环流}) = \frac{4\pi}{c}(\text{电流密度向量过}\ S_2\ \text{的通量}) = 0.$$

如果在 $S_2$ 上也指定 $\boldsymbol{n}$ 为外法线(图 7 - 6 - 2 上的虚线箭头),$S_2$ 在 $L$ 上诱导一个相反的定向(也用虚线箭头表示),$S_1$ 与 $S_2$ 共同包围了一个区域 $\Omega$,而在单位时间内流入 $\Omega$ 的电流是上述两个通量之和,且不会为 0. 于是,随着时间积累,在 $\Omega$ 内的总电荷将达到 $\pm\infty$,这当然是不对的. 如果纯粹从数学观点来看,$\boldsymbol{B}$ 在 $L$ 上的环流由一个 1 - 微分形式 $\boldsymbol{B}\cdot\mathbf{d}\boldsymbol{s} = B_1\mathrm{d}x + B_2\mathrm{d}y + B_3\mathrm{d}z$ 决定. 它在 $L = \partial S_2$ 上积分为 0,表明它属于 0 - 上同调类,而在 $L = \partial S_1$ 上不为 0,则表明它属于另一个非 0 的上同调类. 这种拓扑性质的矛盾是很深刻的. M 本人就已预见到,电磁理论将涉及许多拓扑学问题. 当然,那时拓扑学还处于极为萌芽的状态. 而且称为"位置几何学". M 引述了早期拓扑学研究者李斯廷(J. B. Listing,高斯的学生,"拓扑学"一词似乎是他开始使用的)的著作,而且强调了莱布尼茨和高斯都已理解这一分支的重要性(同上,18 页),M 的著作中对势的多值性、线积分的周期以及多连通区域的刻画的研究,其深刻甚至超过了我们现在通用的教材. 读了 M 的著作更有助于理解 de Rham 上同调理论.

可是更加值得注意的是,所谓电流即电荷的运动. 在某个参考系中是运动的电荷即电流,在另一个参考系中则是静止的电荷. 所以在某一个参考系中只有电场存在,而在另一个参考系中则也有磁场存在. 所以电场和磁场应该看成是一个统一的场的两个侧面,要看我们在一个什么样的参考系中去研究它. 电磁学中涉及运动着的参考系的问题还很多,其最基本的一个是两个电荷之间的力,读者们一定知道这里有著名的库伦定律,这个定律指出,两个静止电荷之间的作用力是

$\boldsymbol{F} = ke_1e_2\boldsymbol{r}/r^3$ (这里取 $k=1$,因为我们采用了高斯——CGS单位制),但是可能不少读者不曾注意到,库伦定律只适用于静止电荷,其理由正在于避开了磁效应.同时适用于电磁场中作用于运动电荷的力的公式则是洛伦兹力的公式:

$$\boldsymbol{F} = q\boldsymbol{E} + \frac{q}{c}\boldsymbol{v} \times \boldsymbol{B}. \tag{3}$$

其实,磁场强度 $\boldsymbol{B}$ 的定义就是(3) 式:磁场强度是一个向量,它规定了作用在一运动电荷上而且与速度成正比的那一部分力.

上面我们说了库伦定律只对静止电荷成立,但是高斯定律即(Ⅰ)则在任何情况下都成立,而且这可以看成是一个实验事实.

选择一定的参考系并且考虑这种选择是否会对物理定律有影响,人们自然会得到以下的结论:当参考系变动时,物理定律应该是协变的.即由一个参考系中该物理定律的数学形式应该能得出任一参考系中该定律的相应形式.这就是一般的相对性原理.但是这个原理之所以有效,却不是由抽象的思辨得出的,而是长期的实验和理论分析的结果.首先从牛顿力学谈起.牛顿力学主要讲惯性系,惯性系就是惯性定律在其中成立的参考系.如果 $Oxyz$ 是一个惯性系而 $O'x'y'z'$ 对它作匀速直线运动,则 $O'x'y'z'$ 也是一个惯性系.在这两个惯性系中一切力学定律形式都是相同的.或者换一个常用的说法,在一个参考系中作任何力学实验均无法确定它是静止的还是作匀速直线运动.这个原理称为狭义相对性原理,因为它只涉及力学实验.现在给它一个数学表述.观测者所在的参考系 $Oxyz$ 称为实验室参考系,他在此参考系中观测某一个物理对象.我们常将另一个参考系 $O'x'y'z'$ 附着在被观测的对象上,并称之为附着参考系.这个物理对象在实验室中运动就说 $O'x'y'z'$ 在 $Oxyz$ 中运动.如果 $O'x'y'z'$ 在 $Oxyz$ 中作匀速直线运动,为简单起见,设这两个参考系的坐标轴平行,而前者沿 $x$ 轴方向以常速 $v$ 运动,我们就会有

$$x = x' - vt', y = y', z = z'; t = t' \tag{4}$$

(见图 7－6－3).(4)称为伽利略变换.于是牛顿力学中的狭义相对应性原理可以表述为:一切力学定律在伽利略变换下是协变的(不变的).若这两个参考系同为惯性系,则它们在力学上是完全等价的,谁也不比谁优越:庄周梦见自己变成了蝴蝶,也可以说是蝴蝶梦见自己变成了庄周.两个说法谁也说服不了谁.因为在此变换中时间是不变的,而空间坐标则发生了变化,所以我们说在牛顿力学中时间仍然是绝对的,但绝对空间不存在了,是在力学中找不到的.在有了电磁场理论以后,人们自然会问:是否任何物理实验(不只是力学实验)均不能在两个惯性系中区别出何者为静止的?如果都不能区别,就说有广义相对性原理成立,M 不是把电磁场看成某种媒质——以太中的场么?如果真是这样,岂不是可以把以太当作一个具有特殊优越性的空间参考系,而宣称在电磁学中找到了绝对空间么?以太是否存在?光速问题给了我们一个线索.

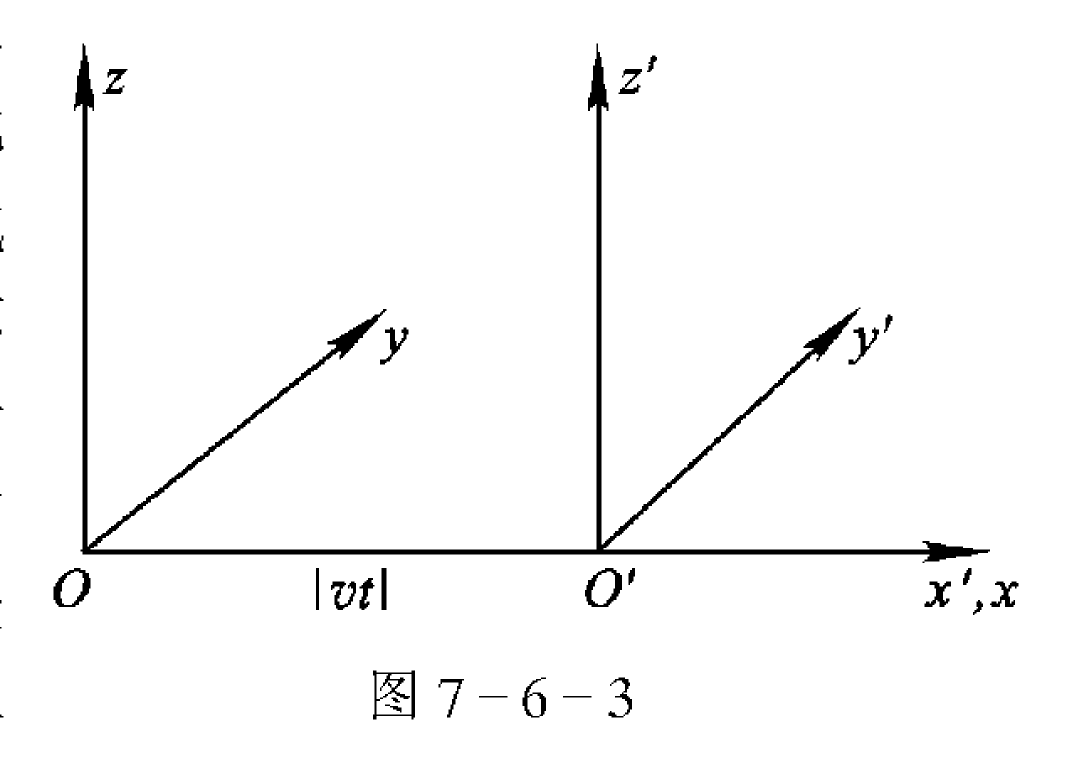


图 7－6－3

M 方程组中的 $c$ 正是光速,这不但 M 早已知道,更早一些韦伯和克希荷夫也从其它角度知道.但一直到1888年赫兹(Heinrich Hertz)发现了由纯粹的电磁过程产生的电磁波具有光的一切性质,才最后明确了光的本性就是电磁波.那么假定以太如一条小河以速度 $v$ 流淌.我们泛舟顺流而下,向小船前进方向射出一束光子,即光线,其速度为 $c$,则岸上的观测者看到的光速应是 $c$

$+v$，然后又溯流而上并向前发出光线. 这时岸上观测者看到的光速应为 $c-v$. 1887 年美国物理学家迈克耳孙（A. A. Michelson）和莫雷（E. W. Morley）按此思想设计了著名的实验. 按当时的实验条件（用光栅测量光波的干涉），应该可以测出光速由 $c+v$ 变到 $c-v$，但是结果是否定的. 这说明以太并不存在，而光速在一切惯性系都相同.

迈克耳孙和莫雷的实验表明，不管在什么样的惯性系中 M 方程组中的光速 $c$ 都是相同的，这就明显地表明 M 方程组没有伽利略不变性. 由于光速不变已是不争的事实，这就必然产生许多悖论性质的推论：在运动方向上，长度会缩短；在运动过程中时钟会变慢；在一个惯性系中同时发生的事件在另一惯性系中并不同时发生，…… 人们纷纷作出各种各样的解释，而爱因斯坦在 1905 年以极大的睿智指出：所有这些"解释"都是不相干的. 所有这些"怪事"其实都是极自然的. 它们来自时空的本性. 牛顿的时空观应该用相对论的时空观来代替：时空并非互相分离的 $\mathbf{R}_x^3$ 与一维的绝对时间之积 $\mathbf{R}_x^3\times\mathbf{R}_t^1$，而是一个 4 维流形 —— 闵可夫斯基时空 $M_4$，它只需要一个局部坐标系 $(x^0,x^1,x^2,x^3)$ 即可刻画，这里 $x^0=ct$. $M_4$ 中也有度量，可是它并不是由一个正定矩阵 $(g_{ij})$，$i.j=0,1,2,3$ 所刻画的黎曼度量，而是由

$$(g_{ij})=\begin{pmatrix}1 & & & 0\\ & -1 & & \\ 0 & & -1 & \\ & & & -1\end{pmatrix} \tag{5}$$

所定义的伪黎曼度量（pseudo-Riemannian metric）. 我们称为洛伦兹度量. 其中两点 $(x^0,x^1,x^2,x^3)$ 与 $(y^0,y^1,y^2,y^3)$ 的距离是

$$I^2(x,y)=(x^0-y^0)^2-\sum_{i=1}^{3}(x^i-y^i)^2 \tag{6}$$

我们应称它为"区间"（interval）. 前面我们讲过的霍奇 $*$ 算子（它在后面是很有用的）中的 $(-1)^{p(n-p)+s}$ 中 $n=4$，$s=3$. 由于广义相对性原理仍要保持，所以伽利略变换必须要代以洛伦兹变换，在最简单的情况下，它是

$$x^0=x'^0\Big/\sqrt{1-v^2/c^2}+\frac{vx'^1}{c}\Big/\sqrt{1-v^2/c^2},$$

$$\left(\text{或 } t=t'\Big/\sqrt{1-v^2/c^2}+\frac{vx'^1}{c^2}\Big/\sqrt{1-v^2/c^2}\right),$$

$$x^1=\frac{vx'^0}{c}\Big/\sqrt{1-v^2/c^2}+x'^1\Big/\sqrt{1-v^2/c^2},$$

$$x^2=x'^2,\qquad x^3=x'^3.$$

可见时间与空间在洛伦兹变换中是紧密地结合在一起的. 这时，广义相对性原理可以表述为：物理定律都应在洛伦兹变换下是协变的（不变的）. M 方程组正是在洛伦兹变换下不变的. 因此我们才说 M 的电磁场理论本质上是相对论性的. 反之，牛顿的力学方程虽然在伽利略变换下是不变的（这里涉及对外力应加什么限制的问题），在洛伦兹变换下却不是不变的，因此需要作相对论修正. 关于洛伦兹变换以及 M 方程组的洛伦兹不变性我们都不加证明了. 它们属于一个极其诱人的领域 —— 相对论，但是我们的读者如果想读一点相对论的话，现在他的数学准备已经大体足够了（关于广义相对论，再读一点协变导数和曲率理论也就够用了）. 下面我们回到 M 的理论.

我们前已指出，现在已经造好了一个新的数学框架，问题是怎样把 M 理论放进去. 新框架是在 $M_4$ 中讨论各种物理问题，因此各种向量都应写成 4 维向量. 关于时空坐标我们已经做到了这一点，即把时间 $t$ 通过 $x^0 = ct$ 当作一个坐标，而得到 $(x^0, x^1, x^2, x^3)$. 关于动量，我们将发现，可以把能量 $E$ 引进去形成 4 维的能量 – 动量向量 $\left(\frac{E}{c}, \boldsymbol{p}\right)$. 力也是一样. 问题是 $\boldsymbol{E}$ 和 $\boldsymbol{B}$ 怎样看成 4 维向量呢？我们在前面已经指出：$\boldsymbol{E}$ 是真向量而 $\boldsymbol{B}$ 是赝向量，二者怎能混在一起呢？而妙不可言的是，正是这两个冤家碰了头，才显示出它们的真实面目：它们都是一个二阶张量的分量的一部分. 4 维流形上的二阶张量有 16 个分量，而 4 维流形上的二阶反对称张量恰好有 6 个分量，恰好与 $\boldsymbol{E}$，$\boldsymbol{B}$ 分量总数相同. 真赝之分正是反对称性的表现. 下面详细道来.

M 本人其实已经开始意识到真赝向量的区别. 他说“还有另一种不同种类的有向量之间的区别；这种区别虽然从物理观点看来是很重要的，但是对数学方法的目的来说却是不必考虑的. 这就是纵向性质和旋转现象之间的区别”. 前者是指那些“完全可以依赖于某一条线而出现的”作用与效应，例如电解质中的电流，后者则是“本性为以该线为轴的转动”的那一种向量，而且 M 明确指出“磁性是一种旋转现象”. 这两种向量服从相同的合成定律，即具有相同的运算法则，所以“对数学方法的目的来说”不必考虑其区别. 显然所谓纵向性质是指真向量，旋转性质是指赝向量. 但是 M 并没有用这个名词，现在也很难追溯是谁先使用这个名词或极向量、轴向量的名词，大约总是 20 世纪初的事. 那时许多数学家和物理学家都在使用向量分析，而四元数就不再为人看好了. 从数学思想看来，认识到这个区别的重要性的人是格拉斯曼. 但是他的思想较晦涩，为人又过于淡泊，不求闻达于诸侯. 何况 M 对哈密尔顿似乎更加情有独钟.

总之，我们现在要在闵可夫斯基时空 $M_4$（其上赋有度量(5)）上找一个二阶反对称协变张量 $F$ 来刻画电磁现象. 我们要找协变张量是因为这样找出的 $F$ 恰好是一个二阶外微分形式，而我们恰好又需要对它施以霍奇 $*$ 算子.

我们来看看法拉第的电磁感应定律. 把一根磁铁推过一个线圈 $L$，必在线圈中产生电流，因而电场强度 $\boldsymbol{E}$ 沿此闭曲线 $L$ 将会作功，也就是产生一个电动势 $U$，它等于 $-\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}t}$（穿过 $L$ 的 $\boldsymbol{B}$ 的通量），这就是(Ⅲ)式：

$$U = \int_L \boldsymbol{E} \cdot \mathbf{d}\boldsymbol{s} = \int_L E_i \mathrm{d}x^i = -\frac{1}{c}\frac{\partial}{\partial t}\Phi. \tag{7}$$

如果作一片曲面 $\Omega$ 使 $L = \partial\Omega$ 而且适当地选择定向，则通量 $\Phi$ 应该是

$$\Phi = \iint_\Omega B_1 \mathrm{d}x^2 \wedge \mathrm{d}x^3 + B_2 \mathrm{d}x^3 \wedge \mathrm{d}x^1 + B_3 \mathrm{d}x^1 \wedge \mathrm{d}x^2.$$

用 $\omega$ 表示 2 阶微分形式

$$\omega = B_i \mathrm{d}x^{i+1} \wedge \mathrm{d}x^{i+2}. \tag{8}$$

$(i, i+1, i+2)$ 是 $(1,2,3)$ 的轮换排列，再令 $\sigma = E_i \mathrm{d}x^i$ 为一个 1 阶微分形式，则由斯托克斯定理，有

$$\int_{\partial\Omega} \sigma = \iint_\Omega \mathrm{d}\sigma = -\frac{1}{c}\iint_\Omega \dot{\omega}.$$

因此我们可把方程(Ⅲ)用外微分形式改写为

$$d\sigma = -\frac{1}{c}\dot{\omega}.$$

可以看到 $\omega$ 应是我们想求的二阶反对称协变张量的“一部分”. 按我们记二阶反对称协变张量的分量 $A_{\mu v}$ 时规定 $\mu < v$，所以应把 $\omega$ 写成

$$\omega = B_1 dx^2 \wedge dx^3 - B_2 dx^1 \wedge dx^3 + B_3 dx^1 \wedge dx^2.$$

另一部分应是 $(\cdot)dx^0 \wedge dx^1 + (\cdot)dx^0 \wedge dx^2 + (\cdot)dx^0 \wedge dx^3$，这些未定的系数当然也可以从物理上推导出来，但因为要利用闵可夫斯基力与“内部乘积”(interior product) 的概念，都是我们未曾讲到的，所以我们只好试一下，令它们为 $E_1, E_2, E_3$ 于是得到

$$F = (E_1 dx^1 + E_2 dx^2 + E_3 dx^3) \wedge dx^0 + B_1 dx^2 \wedge dx^3 - B_2 dx^1 \wedge dx^3 + B_3 dx^1 \wedge dx^2. \tag{9}$$

现在看一下 $dF = 0$ 究竟是什么. 它是一个3阶外微分，把它算出来，并注意到例如 $\frac{\partial B_1}{\partial x^0} = \frac{1}{c}\frac{\partial B_1}{\partial t}$，即有

$$\begin{aligned} dF = & \left(\frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3}\right) dx^0 \wedge dx^2 \wedge dx^3 + \left(\frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2}\right) dx^0 \wedge dx^1 \wedge dx^2 \\ & + \left(\frac{\partial E_3}{\partial x^1} - \frac{\partial E_1}{\partial x^3}\right) dx^0 \wedge dx^1 \wedge dx^3 + \frac{1}{c}\frac{\partial B_1}{\partial t} dx^0 \wedge dx^2 \wedge dx^3 \\ & - \frac{1}{c}\frac{\partial B_2}{\partial t} dx^0 \wedge dx^1 \wedge dx^3 + \frac{1}{c}\frac{\partial B_3}{\partial t} dx^0 \wedge dx^1 \wedge dx^2 \\ & + \left(\frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3}\right) dx^1 \wedge dx^2 \wedge dx^3. \end{aligned}$$

所以 $dF = 0$ 就化成了 M 方程组中的(Ⅲ)与(Ⅱ)

$$\textbf{curl } \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t}, \quad \text{div } \boldsymbol{B} = 0.$$

可见 $F$ 的选法是正确的.

现在看一下，这样选择的 $F$ 怎样把真向量 $\boldsymbol{E}$ 与赝向量 $\boldsymbol{B}$ 结合起来. 为此改到左手坐标系，这就相当于把 $x^2$ 与 $x^3$ 对调一下，于是 $F$ 就变成了

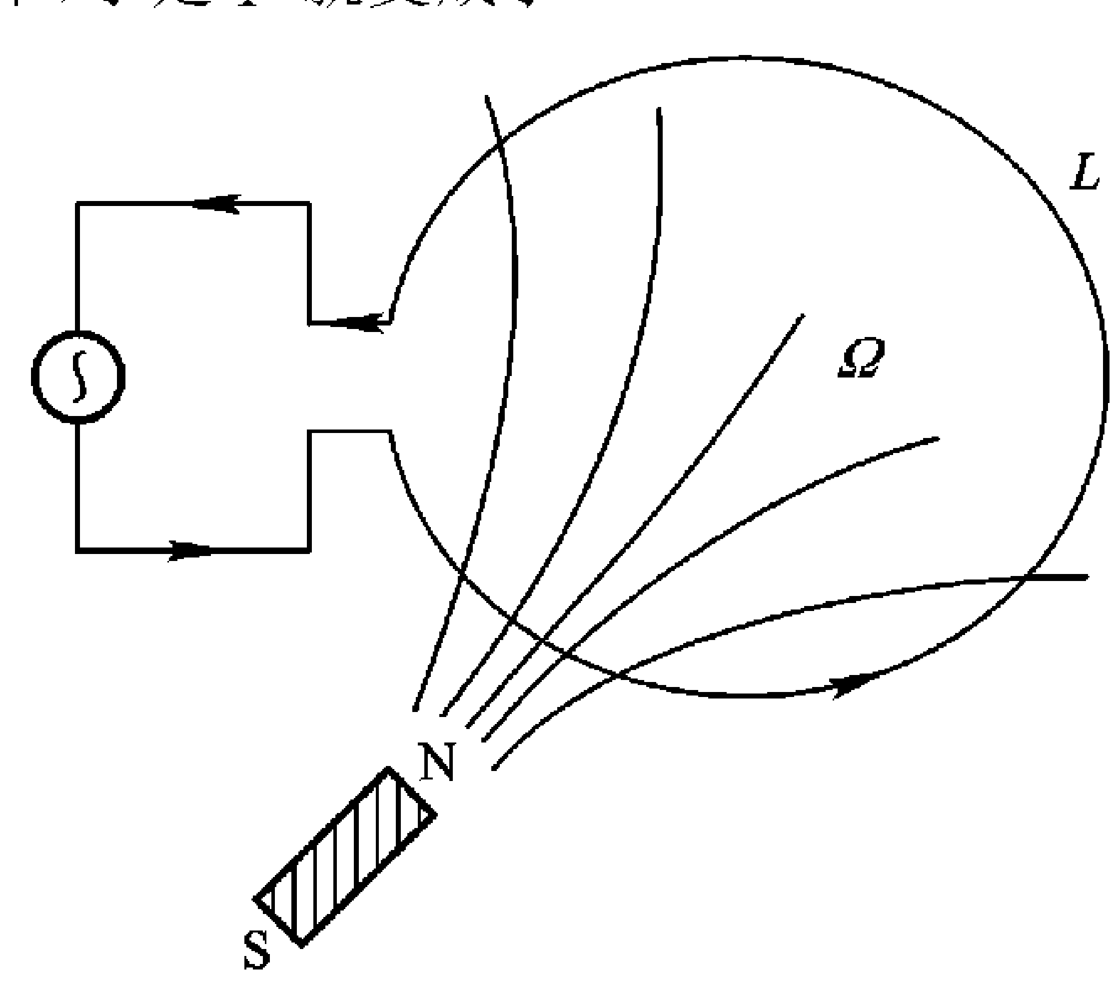



图 7-6-4

$$\tilde{F} = (E_1\mathrm{d}x^1 + E_2\mathrm{d}x^3 + E_3\mathrm{d}x^2) \wedge \mathrm{d}x^0 + B_1\mathrm{d}x^3 \wedge \mathrm{d}x^2 - B_2\mathrm{d}x^1 \wedge \mathrm{d}x^2 + B_3\mathrm{d}x^1 \wedge \mathrm{d}x^3$$
$$= (E_1\mathrm{d}x^1 + E_3\mathrm{d}x^2 + E_2\mathrm{d}x^3) \wedge \mathrm{d}x^0 - B_1\mathrm{d}x^2 \wedge \mathrm{d}x^3 - B_2\mathrm{d}x^1 \wedge \mathrm{d}x^2 + B_3\mathrm{d}x^1 \wedge \mathrm{d}x^3.$$

如果仍把它写为

$$\tilde{F} = (\tilde{E}_1\mathrm{d}x^1 + \tilde{E}_2\mathrm{d}x^2 + \tilde{E}_3\mathrm{d}x^3) \wedge \mathrm{d}x^0 + \tilde{B}_1\mathrm{d}x^2 \wedge \mathrm{d}x^3 - \tilde{B}_2\mathrm{d}x^1 \wedge \mathrm{d}x^3 + \tilde{B}_3\mathrm{d}x^1 \wedge \mathrm{d}x^2.$$

就看出,电场强度变成了

$$(E_1, E_2, E_3) \to (\tilde{E}_1, \tilde{E}_2, \tilde{E}_3) = (E_1, E_3, E_2):$$

这只把第二、三分量对换,所以是真向量;但是磁场的变化却是

$$(B_1, B_2, B_3) \to (\tilde{B}_1, \tilde{B}_2, \tilde{B}_3) = -(B_1, B_3, B_2).$$

不但二、三分量对换,而且全部变了符号. 所以是赝向量. 总之,三维向量的真赝之分只不过是 4 维张量反对称性的表现.

要想得到 M 方程组的另外两个,我们先来计算 $F$ 之对偶张量 $*F$. 注意. $F$ 是 2 阶反对称协变张量,$p = 2$,流形 $M_4$ 的维数 $n = 4$,而度量中负特征根个数 $s = 3$,所以 $(-1)^{p(n-p)+s} = -1$. 这样

$$-*F = E_1\mathrm{d}x^2 \wedge \mathrm{d}x^3 + E_2\mathrm{d}x^3 \wedge \mathrm{d}x^1 + E_3\mathrm{d}x^1 \wedge \mathrm{d}x^2 - B_1\mathrm{d}x^0 \wedge \mathrm{d}x^1 + B_2\mathrm{d}x^0 \wedge \mathrm{d}x^2 + B_3\mathrm{d}x^0 \wedge \mathrm{d}x^3.$$

所以

$$-\mathrm{d}*F = \left(\frac{\partial E_1}{\partial x^1} + \frac{\partial E_2}{\partial x^2} + \frac{\partial E_3}{\partial x^3}\right)\mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 + \left(\frac{1}{c}\frac{\partial E_1}{\partial t} + \frac{\partial B_2}{\partial x^3} - \frac{\partial B_3}{\partial x^2}\right)\mathrm{d}x^0 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3$$
$$+ \left(-\frac{1}{c}\frac{\partial E_2}{\partial t} + \frac{\partial B_1}{\partial x^3} - \frac{\partial B_3}{\partial x^1}\right)\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^3 + \left(\frac{1}{c}\frac{\partial E_3}{\partial t} + \frac{\partial B_1}{\partial x^2} - \frac{\partial B_2}{\partial x^1}\right)\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2.$$

现在我们想把(Ⅰ)和(Ⅱ)代进去. (2) 和(Ⅳ)的右方形成一个 4 维向量 $4\pi\left(\rho, \frac{1}{c}j^1, \frac{1}{c}j^2, \frac{1}{c}j^3\right)$. 我们先用 $M_4$ 中的度量 (5) 相应的 $(g^{ij})$,$i, j = 0,1,2,3$ 去作用在这个 4 维向量上. 但因 $(g^{ij}) = (g_{ij})^{-1} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} = (g_{ij})$,所以于是得到一个 4 维向量:$J = \left(\rho, -\frac{1}{c}j^1, -\frac{1}{c}j^2, -\frac{1}{c}j^3\right) = \rho\mathrm{d}x^0 - \frac{1}{c}j^1\mathrm{d}x^1 - \frac{1}{c}j^2\mathrm{d}x^2 - \frac{1}{c}j^3\mathrm{d}x^3$. 如果要换到其他坐标系,则把 $J$ 当作一个协变向量来处理. 于是 $J$ 成了一个协变的 4 维向量,称为电荷 – 电流向量. 现在对它施用霍奇算子 $*$ 把它化为 3 微分形式. 利用 §3 中关于计算 $*\omega$ 的公式于此,有

$$*J = -\rho\mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 + \frac{j^1}{c}\mathrm{d}x^0 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 - \frac{j^2}{c}\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^3 + \frac{j^3}{c}\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2.$$

把它与 $\mathrm{d}*F$ 比较,立即得出

$$\mathrm{d}*F = 4\pi *J. \tag{10}$$

如果再把它化为 $\delta = - * \mathrm{d} *$，利用 $* * \omega = (-1)^{p(n-p)+s}\omega$，这里 $p = 1, n = 4, s = 3$，上式就可以化为

$$\delta F = 4\pi J.$$

总之，M 方程组现在化成了关于二阶反对称协变张量 $F$ 的两个形状极为简洁的方程

$$\mathrm{d}F = 0, \tag{MI}$$

$$\delta F = 4\pi J. \tag{MⅡ}$$

$F$ 称为电磁强度张量，或称为法拉第张量，它的矩阵形式是

$$F = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}.$$

把 M 方程组写成这样简洁的形式，就可以很方便地把本章讲的数学工具应用于它. 作为一个例子，把算子 d 作用于(10) 式两侧，因为 $\mathrm{d}^2 = 0$，所以就有 $4\pi \mathrm{d} * J = 0$，把 $* J$ 的表达式代入，立即有连续性方程

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \boldsymbol{j} = 0. \tag{11}$$

可见，连续性方程即电荷守恒定律，其实是 M 方程组的推论. 我们这里未能介绍的关于电磁场的能量关系，也都可以由此导出. 这样我们就可以说，M 方程组确实是整个电磁理论的基础. 但是关于作用在电荷上的力的洛伦兹力的公式却不能由 M 方程组导出. 因为一则其中出现了电荷的运动速度 $\boldsymbol{v}$ 不曾在 M 方程组中出现，而且也不可能出现，因为 M 方程组是洛伦兹不变的，不论电荷以什么速度运动，只要时间与空间的坐标的变化适合洛伦兹变换，M 方程组总是不变的. 有了力，当然就会考虑牛顿力学如何应用于它. 但是 M 理论本质上是相对论性质的，而牛顿力学则不是. 因此，要想把牛顿力学应用于电荷的运动，就应该加上相对论修正.

此外，M 方程组虽然形状写起来十分简单，其内容仍是复杂的，要去求解就很不容易. 不过我们注意到方程(MI)，由庞加莱引理就知道，在 $M_4$ 的任一个星形域中一定存在一个一阶微分形式 $A = \varphi \mathrm{d}x^0 + A_i \mathrm{d}x^i$ 使

$$F = \mathrm{d}A. \tag{12}$$

$A$ 称为势，$\varphi$ 是它的时间部分，它就是经典的物理教本中的标量势，其余的三个分量构成其空间部分，我们用古典的向量分析记号(即黑体字母) 来表示它：$\boldsymbol{A} = (A_1, A_2, A_3)$，它就是通常所谓向量势. 于是

$$\begin{aligned} \mathrm{d}A &= \sum_{\mu<\nu} \left(\frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A_\nu}{\partial x_\mu}\right) \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu, \mu, \nu = 0,1,2,3 \\ &= \sum_{i=1}^{3} \left(\frac{\partial A_i}{\partial x^0} - \frac{\partial \varphi}{\partial x^i}\right) \mathrm{d}x^0 \wedge \mathrm{d}x^i \\ &\quad + \sum_{i<j} \left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}\right). \end{aligned}$$

与 $F$ 的表达式(9) 比较，即得

$$E = -\textbf{grad}\ \varphi - \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A}. \tag{13}$$

$$\boldsymbol{B} = \textbf{curl}\ \boldsymbol{A}. \tag{14}$$

这与传统的讲法完全一致.

但是同一个电磁场 $F$ 可以有许多不同的势.因为若 $A$ 是一个势,加上一个0阶微分形式 $\lambda$ 的外微分,即 $M_4$ 上的一个光滑函数 $\lambda$ 的全微分 $\mathrm{d}\lambda$,则 $A' = A + \mathrm{d}\lambda$ 也是一个势,因为

$$\mathrm{d}A' = \mathrm{d}A + \mathrm{d}^2\lambda = \mathrm{d}A = F.$$

反过来,若 $A'$ 是 $F$ 的另一个势,则 $\mathrm{d}(A' - A) = 0$.而由庞加莱引理,一定存一个0阶微分形式,即一个光滑函数 $\lambda$ 使 $A' = A + \mathrm{d}\lambda$.这里当然都是在 $M_4$ 的一个星形区域中讲的.总之 $A + \mathrm{d}\lambda$ 给出了全部的势,任取一个都可以得到同样的 $F$.这个性质称为 $F$ 的规范不变性.这样的说法来由如下:在原点处一个单位电荷必在 $(x, y, z)$ 点产生电势 $\frac{1}{r} + C$,通常我们取 $C = 0$,即要求势在无穷远点为0,这就是一种规范.现在讲的则与此类似.物理学中有许多场都有规范不变性,即所谓规范场.我们所讲的只是最简单的情况.

既然有多种不同的势,我们必可取其中一个使所讨论的方程最简单,这就是要求 $\delta A = 0$.因为 $\delta A = \pm * \mathrm{d} * A$,而

$$\pm * A = \varphi \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 - A_1 \mathrm{d}x^0 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 + A_2 \mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^3 - A_3 \mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2.$$

所以

$$\pm \delta A = \frac{1}{c}\frac{\partial \varphi}{\partial t} + \mathrm{div}\ \boldsymbol{A} = 0. \tag{15}$$

这里我们采用记号 $\pm$ 是为了在计算 $\delta = * \mathrm{d} *$ 时更方便一些.规范(15)有特殊的重要性,称为 Lorenz 规范(请注意这不是爱因斯坦相对论的伟大先行者洛伦兹(H. A. Lorentz)而是另一个人 L. Lorenz,尽管拼法与发音都很相近).这种规范必满足方程

$$\Delta A = (\delta \mathrm{d} + \mathrm{d}\delta) A = \delta F = 4\pi J. \tag{16}$$

不过这里的 Laplace-Beltrami 算子并非古典的拉普拉斯算子 $\sum_{i=0}^{3}\left(\frac{\partial}{\partial x^i}\right)^2$,而是波算子 $\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \sum_{i=1}^{3}\left(\frac{\partial}{\partial x^2}\right)^2 = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta_3$,而且时常记成 $\square$,称为达朗贝尔算子(d'Alembert operator).因为现在的度量不是欧几里得度量,而是洛伦兹度量.这里我们没有详细推算.但是不论如何,它是 Laplace-Beltrami 算子,而前面指出,它作用在1阶微分形式 $A$ 上时,可以分别作用到各个分量上去.因此有

$$\square \varphi = 4\pi\rho.$$

$$\square \boldsymbol{A} = 4\pi \boldsymbol{j}.$$

现在我们把势的时间部分与空间部分完全分开了.电荷决定 $\varphi$(而不是如我们习惯想像的:电荷决定电场),电流决定 $\boldsymbol{A}$(而不是如我们习惯想像的:电流决定磁场).

这里的方程是著名的波动方程.它是偏微分方程中极广泛的一个分支.由此我们看到了,打开了波动方程以及更一般的双曲型方程应用于电磁波理论的道路.

可是我们把 M 方程组化为(MⅠ)和(MⅡ)还有更重要的意义:它虽然比较抽象,但却是一

种极重要的框架.再加上这个框架的进一步发展,其中可以包含许多更深刻反映大自然本质的场,在 21 世纪中我们必可看到它的大发展.这当然绝不是微积分的创始者们,如 17 世纪的牛顿所能梦想得到的了.